A. S. KOMPANEYETS

A COURSE OF THEORETICAL PHYSICS

VOLUME 1

FUNDAMENTAL LAWS

MECHANICS
ELECTRODYNAMICS
QUANTUM MECHANICS

A. S. KOMPANEYETS

A COURSE OF THEORETICAL PHYSICS

VOLUME 1

A. S. KOMPANEYETS

A COURSE OF THEORETICAL PHYSICS

PHYSICS VOLUME 1

A. S. KOMPANEYETS

FUNDAMENTAL LAWS

Translated from the Russian by V. TALMY

MIR PUBLISHERS
MOSCOW

А. С. КОМПАНЕЕЦ

КУРС ТЕОРЕТИЧЕСКОЙ ФИЗИКИ

ТОМ 1
ЭЛЕМЕНТАРНЫЕ ЗАКОНЫ

PREFACE

Theoretical physics, though most intimately connected with experimental physics, differs substantially from the latter both in methods and results.

Experiments establish individual facts, some of them of outstanding, fundamental significance. Theory explains them, but it also formulates general principles. If the task of theoretical physics were merely to analyse the results of experiments, it would be no more than the experimenter's aid.

Such theoretical achievements as the enunciation of mechanics, electrodynamics, statistical and quantum mechanics go far beyond simply interpreting a number of separate experiments. Theoretical physics has in many ways contributed to the moulding of mankind's scientific world outlook as a whole, it has influenced the thinking of people far removed from the natural sciences.

Knowledge of theoretical physics requires knowledge of experimental physics, it requires an understanding of the relationships between physical phenomena and general laws; in short, everything one could call knowledge of physics in general is both important and useful.

But equally important is the mastery of the main tool of theoretical physics, mathematics; whatever branch of mathematics one invokes, one must fully master its basic idea and general method. As Lev Landau remarked half-jokingly, it is possible to be a theoretical physicist without really knowing physics, but not without knowing mathematics.

The sheer volume of essential mathematical knowledge is the main obstacle in the study of theoretical physics. Add to this that textbooks on mathematics set forth the material as the mathematicians see right, which is not how theoretical physicists need it.

This book attempts to present theoretical physics in a manner that would require the student to have the barest reasonable minimum of advance mathematical knowledge: elements of infinitesimal calculus, the beginnings of analytical geometry, and vector algebra. The course sets forth basic information in vector analysis, matrix, tensor and spinor algebra, and a small section devoted to spherical functions. These mathematical explications are in part incorporated in the main text, in part in the exercises, where the respective problems are provided with worked solutions, which

differ from the study material in the conciseness of their presenta-
tion. Many physical problems are also provided with such solutions.
The best way of setting forth study material is always open to debate.
If one attempts to sidestep complex or subtle questions or oversimpli-
fy the argumentation in the hope that the unsophisticated reader
will fail to notice a certain "sleight of hand", the best that can be
achieved is merely an illusion of understanding, suitable only in the
reading of popular science books.

On the other hand, a text overburdened with clarifications and
reservations, with lengthy discourses on every issue and what is
known in mathematics as "epsilontics", is capable of confusing the
student and obscuring the essentials behind trivialities.

The art of the theoretician as a teacher consists in an ability to
find the optimum mode of presentation, guided by his prospective
audience or readership.

In order to gain a real understanding of theoretical physics it is
essential to constantly bear in mind the fundamental ideas of the
subject, the purpose of the specific discourse or computation, and
the connections among all the details and general principles. Further-
more, in theoretical physics there is no knowledge without practical
skill: only he understands the subject who has properly mastered
its methods. Passive digestion is impossible here, simply memoriz-
ing is useless.

This course consists of two volumes. The first sets forth the funda-
mental laws of physics. The second, the statistical laws of large
assemblies of particles—gases, liquids, solids. The laws that emerge
in such assemblies are based on the laws of elementary interactions
and the properties of large numbers.

The presentation of the subject matter adopted here appears to
be most suitable for the purposes of this book. There is virtually
no information of a purely abstract character, containing only
summaries of results or mentioning new ideas without expounding
upon them. Whatever is asserted is deduced from general laws.
In each chapter these laws precede everything else: it is hardly ex-
pedient to set forth basic principles after all the specifics are already
known.

I have attempted, as far as possible, to set forth the material in
my own way. However, I have borrowed freely from the many-
volumed encyclopedic *Course of Theoretical Physics* by L.D. Landau
and E.M. Lifshitz, as well as from *The Principles of Quantum Me-
chanics* by P.A.M. Dirac.

A.S. Kompaneyets

CONTENTS

PART I

MECHANICS

1

GENERAL REMARKS

Mechanics is the bedrock of all theoretical physics. Starting out with certain introductory propositions, we shall present the equations of mechanics in the form most convenient for solving the concrete problems of dynamics and allowing for generalization over other domains of physical theory.

Frames of Reference. To describe the motion of a mechanical system it is necessary to specify its position in space as a function of time. It is meaningful to speak only of the relative position of any body. For example, the position of a ship at sea is given by its latitude and longitude relative to specified points and lines on the globe. The position of a body in space can be measured relative to the sun or the centre of the Galaxy, etc.

Besides stating the coordinates of the bodies of a system relative to a selected coordinate system it is necessary to specify the time at which the coordinates assume the given values. In other words, one needs a clock. Usually this is a uniform periodic process, natural or artificially reproduced in a mechanism, such as the rotation of the earth around its axis.

The intuitive concept of a single universal time to which we are accustomed in everyday life is true only when the relative velocities of all bodies are small in comparison with the velocity of light. It is in the framework of this approximation that Newtonian mechanics is valid. In the more general case it is necessary to state the system of bodies in which the clock used for measuring the time is fixed. If the coordinates stating the positions of moving bodies are

fixed within the same system, it is said that a definite *frame of reference* has been defined. (Thus, the position of a ship at a specific time is verified against a clock located somewhere on the shore.) In Newtonian mechanics it is assumed that the readings of clocks are the same in all reference frames. We should remember, however, that this is a property not of time in general but only of reference systems of bodies moving slowly relative to one another. Here we shall accept this approximation. The generalizations will be made in Part II.

Newton's Second Law. Motion in mechanics consists in changes in the mutual spatial configurations of bodies with time, relative to a certain selected frame of reference. In formulating the laws of motion an extremely convenient concept is that of a *mass point*, or *particle*, that is, a body whose position in space is fully defined by three Cartesian coordinates. Strictly speaking, this idealization is inapplicable to any real body. Nevertheless, it is quite reasonable when a body's motion is sufficiently well defined by the displacement of any of its points and is independent of its rotations or deformations.

The earth's motion around the sun is independent of its rotation around its axis, while the flight of a bullet is strongly dependent on its rotational motion. In this sense the earth approximates the concept of a mass point more closely than a bullet. The absolute dimensions of bodies are immaterial.

If we proceed from the concept of a mass point as the fundamental entity of mechanics, the law of motion (*Newton's Second Law*) is formulated thus:

$$m \frac{d^2\mathbf{r}}{dt^2} = \mathbf{F} \tag{1.1}$$

Here, $\mathbf{F}$ is the resultant of all the forces applied to the particle (the vector sum of the forces) and $\frac{d^2\mathbf{r}}{dt^2}$ is the acceleration vector the Cartesian components of which are

$$\frac{d^2x}{dt^2}, \qquad \frac{d^2y}{dt^2}, \qquad \frac{d^2z}{dt^2}$$

The quantity m involved in the equation characterizes the mass point and is called its *mass*.

Force and Mass. Equation (1.1) is the physical definition of force. It should not, however, be seen as a simple identity or designation, because underlying it are a number of assumptions concerning the laws of motion. These assumptions are confirmed by the totality of

experimental data regarding mechanical motion. For example, the fact that Eq. (1.1) involves a second derivative with respect to time means that its solution requires a statement of the initial values of the coordinates and velocities, that is, the first derivatives, which is sufficient to determine the coordinates at all subsequent instants. Obviously, this fact can be deduced only from experimental data.

Equation (1.1) essentially defines the mode of interaction between bodies, indicating that it is effected in the form of forces imparting acceleration. This assertion, too, obviously derives from a generalization of experimental data.

Newtonian mechanics makes a limiting assumption regarding force. It is assumed to depend only upon the mutual configuration of the bodies at the instant to which the equation refers, and to be explicitly independent of their configurations at preceding instants. As we shall see later in Part II, this assumption is valid only when the velocities of the bodies are small in comparison with the velocity of light. At large velocities the very definition of interaction changes substantially and cannot be expressed in the form (1.1).

Equation (1.1) involves the quantity m characterizing the body, its mass. The mass of different bodies can be compared according to the acceleration which one and the same force imparts to them. The greater a body's acceleration the less its mass. The mass of some body may be chosen as a standard, the choice being quite independent of the standards of length and time. The dimensions, or unit of measurement, of mass is thus a special one, unrelated to the units of length and time.

The properties of mass are established experimentally. Firstly, when two identical bodies are joined together, the result is a body of double mass in comparison with either of them. A duly stretched spring imparts to such a composite body half the acceleration it would to either of its components. In other words, mass is an additive quantity, as is said when any quantity characterizing a body as a whole equals the sum of those quantities for all its parts separately. Experience indicates that the principle of additivity of mass is also applicable to bodies made up of different substances.

Note that the widespread definition of mass as quantity of matter is meaningless since it does not state the manner in which the quantity is measured. The definition of mass from Eq. (1.1), on the other hand, contains such a statement.

In Newtonian mechanics the mass of a body is a constant quantity which does not change in the body's motion. But the additivity and constancy of mass follow solely from experimental data and is in no way self-evident. These data are restricted to a specific domain of phenomena, namely those in which the forces of interaction do not accelerate bodies to speeds comparable with the velocity of light.

In other words, the interactions should not, in a way, be too strong. In atomic nuclei, where interactions are strong, the principle of the additivity of mass holds to an accuracy of only fractions of a percentage point.

We may note that if, instead of subjecting a body to the force of a stretched spring, it is subjected to the action of gravity, the acceleration of a body of double mass is equal to the acceleration of each mass separately. It must be concluded from this that the force of gravity is, for some reason, itself proportional to the mass of a body. It is this exceptional property of gravity that lies at the basis of Einstein's theory of gravitation.

Inertial Frames of Reference. Equation (1.1) involves the acceleration of a particle. It is meaningless to speak of acceleration without stating the frame of reference in which it is measured. It is therefore necessary to establish what, in each specific case, causes the acceleration. In other words, one must determine whether the acceleration is due to interactions between bodies or to the motion of the reference frame itself. For example, the jolt felt by a passenger when a train brakes suddenly is evidence of the train's nonuniform motion relative to the earth. No one on the platform feels the jolt. This means that the passenger's acceleration cannot be ascribed to interaction forces. Thus, a reference frame fixed with respect to the earth is distinguished by the property that in it accelerations of bodies are due solely to their interactions, for example, to the action of the force of gravity. Finer effects associated with the rotation of the earth will be discussed elsewhere.

It can be supposed that there exists an ideal frame of reference in which all accelerations of bodies are due solely to forces of interaction. Obviously, a frame connected with the earth approximates such an ideal system more closely than one connected with the train.

Whether a given force is due to interaction between bodies can be determined with the help of *Newton's Third Law*: such forces are equal in magnitude and opposite in sense for any pair of interacting particles. This is valid only if the forces are transmitted instantaneously, an assumption which Newtonian mechanics permits.

If the acceleration of bodies in a given reference frame is due solely to their interactions, such a reference frame is called *inertial*. In an inertial frame of reference, a free material point not subject to the action of any other bodies moves uniformly in a straight line.

The direction of gravity on the surface of the earth is determined with the help of a plumb line. But a rock dropped from a tall tower does not fall directly along the plumb line: it is deflected slightly to the east. Consequently this acceleration component is not due to the attraction of the earth, which proves the noninertiality of a reference frame connected with it.

Ideal Constraints. Bodies in contact give rise to forces of interaction which can be described with the help of the kinematic concept of *ideal constraints*. Such constraints cause the points of a mechanical system to move along definite surfaces. If, for example, two material points are joined by an ideal inextensible constraint which holds them at a constant distance, any one of them moves along a sphere with the other point at its centre.

In the most general case the restrictions imposed by constraints cause bodies to move in curved paths. Such motion is always accelerated. The acceleration can be formally ascribed to forces called the *reaction forces of the ideal constraints*. These forces are not given in advance as functions of the positions of the points. Integration of equations of type (1.1) for the case of additional restrictions imposed by constraints gives the reaction forces. In the next section we shall consider a method whereby the reaction forces can be bypassed in solving equations of motion.

Besides reaction forces, motion along a rigid surface also leads to the appearance of friction forces. Their importance in applied mechanics is extremely great. But in motion with friction motion is imparted not only to the body as a whole but to its component molecules as well. The interaction between surfaces slipping over one another is extremely complex and acquires the form of a certain force of interaction only as a result of averaging over individual molecules. In this part we consider the fundamental laws referring to separate material points (particles), not to large associations of molecules. Friction forces are, accordingly, ignored. They are studied in detail in courses on theoretical mechanics.

2

LAGRANGE EQUATIONS

Equation (1.1) was written in Cartesian coordinates. But any coordinate system is a matter of free choice, which means that when we describe some natural law in it we introduce an element of arbitrariness. Furthermore, we are also free to adopt a reference frame of our choice. The velocities of mass points relative to different reference frames are different. But it is desirable to formulate natural laws in such a way as to exclude, as far as possible, quantities which by definition refer to the observer (for example, coordinates) or, in other words, to exclude the element of arbitrariness from the description.

For this we must pass from the differential law (1.1) to an integral one. The value of an integral does not depend on the variables in which it is computed (for example, the area of a figure is the same whatever coordinates are used to calculate it: rectangular, polar, etc.). We can therefore hope to formulate the laws of mechanical motion in a way that would reduce them to statements involving integral expressions describing a certain finite segment of the motion.

This is possible under the following conditions in mind:

(i) The constraints are ideal, that is, there are no friction forces.

(ii) The interaction between mass points can be represented in the form

$$\mathbf{F}_i = -\frac{\partial U}{\partial \mathbf{r}_i} \tag{2.1}$$

where the subscript i refers to a mass point, and the vector quantity $\mathbf{r}_i$ represents a vector with components $\partial U/\partial x_i$, $\partial U/\partial y_i$, $\partial U/\partial z_i$.[1] The quantity U is the same for the whole of the mechanical system. Its meaning will be discussed later.

Hamilton's Principle. Condition (2.1) is not as restrictive as it might appear. It holds for gravity forces, electrostatic forces, elastic forces, that is, precisely those to which Newtonian mechanics apply. From now on we shall express forces in the form (2.1). For the sake of simplicity, in subsequent formulas we shall assume that there is only one ideal constraint. This restriction is of no great significance since the transition to the case of several constraints is performed directly. We write the constraint condition in the form of the equation

$$F (\mathbf{r}_1, \ldots, \mathbf{r}_i, \ldots) = 0 \tag{2.2}$$

Now consider a certain change in the coordinates of the mass points of the system, $\delta \mathbf{r}_i$, which we assume to be infinitesimal. The change is not due to the motion of the points and can be treated as a purely speculative operation. It should not, however, violate condition (2.2), that is, it should be compatible with the constraints imposed upon the system. For example, if the points are compelled to move along a surface, the changes $\delta \mathbf{r}_i$ are taken along the surface, while being completely arbitrary in other respects. On the other hand, if as a result of the displacements the points remain on the surface defined by Eq. (2.2), the displacements satisfy the obvious condition

$$F (\ldots \mathbf{r}_i + \delta \mathbf{r}_i \ldots) - F (\ldots \mathbf{r}_i \ldots) = \sum_i \frac{\partial F}{\partial \mathbf{r}_i} \delta \mathbf{r}_i = 0 \tag{2.3}$$

---

[1] More on differentiation with respect to a vector in the introduction to Part II.

Here use was made of the fact that the quantities $\delta\mathbf{r}_i$ are infinitesimal, so that F expands in a Taylor series up to first derivatives inclusive.

Consider a set of differential equations (1.1) with the supplementary condition (2.2). This condition means that not all the variables $\mathbf{r}_i$ are independent. To make the number of independent variables equal to the number of equations we multiply each equation by the corresponding quantity $\delta\mathbf{r}_i$ and add them. We resolve force $\mathbf{F}_i$ into two components: $\mathbf{F}_i = -\partial U/\mathbf{r}_i + \mathbf{F}'_i$. The first component is due to the interaction between the material points, the second describes the forces due to the action of the constraints.

We now make use of the condition that the constraints are ideal. We start with the simplest case of a smooth unchangeable surface along which a mass point moves. The reaction force is perpendicular to the surface, that is, the scalar product of vectors $\mathbf{F}'_i$ and $\delta\mathbf{r}_i$ is zero, defining, as it does, work done in the displacement of the mass point along the surface, that is, the work of friction forces, which we excluded in advance when we assumed the constraint to be ideal.

In the case of two or, in general, several mass points, the ith components $\mathbf{F}'_i\delta\mathbf{r}_i$ need not all separately vanish, since the points may perform work on one another. For example, if two points are joined by an ideal inextensible constraint and one of them is in some way accelerated, it will draw the other point after it, that is, perform work on it. Thus, in a system of several mass points joined by ideal constraints the condition

$$\sum_i \mathbf{F}'_i\delta\mathbf{r}_i = 0 \tag{2.4}$$

is imposed on the reaction forces, the displacements being subject to Eq. (2.3).

But it then follows from Eqs. (1.1) and (2.4) that the equation

$$\sum_i \left(m_i \frac{d^2\mathbf{r}_i}{dt^2} + \frac{\partial U}{\partial\mathbf{r}_i} \right) \delta\mathbf{r}_i = 0 \tag{2.5}$$

should hold for all displacements compatible with the constraints, that is, satisfying Eq. (2.3). One of the displacements $\delta\mathbf{r}_i$ can be excluded from the latter and substituted into (2.5), after which all the other displacements, obviously, become independent.

It is more convenient to use the *method of undetermined multipliers*, since it becomes possible to preserve the symmetry of the formulas with respect to all $\delta\mathbf{r}_i$'s. Multiply Eq. (2.3) by a factor α and add the result to (2.5) to get

$$\sum_i \left(m_i \frac{d^2\mathbf{r}_i}{dt^2} + \frac{\partial U}{\partial\mathbf{r}_i} + \alpha\frac{\partial F}{\partial\mathbf{r}_i} \right) \delta\mathbf{r}_i = 0 \tag{2.6a}$$

Since α is arbitrary, we have introduced an extra parameter into the equation, thanks to which we may consider all the displacements to be quite independent of one another. It is therefore possible to put all $\delta\mathbf{r}_k$ $(k \neq i)$ equal to zero, except for one, $\delta\mathbf{r}_i$. Then there remains

$$\left(m_i \frac{d^2\mathbf{r}_i}{dt^2} + \frac{\partial U}{\partial \mathbf{r}_i} + \alpha \frac{\partial F}{\partial \mathbf{r}_i} \right) \delta\mathbf{r}_i = 0 \tag{2.6b}$$

Thanks to the parameter α, no binding conditions are imposed upon $\delta\mathbf{r}_i$ anymore, and it can now be treated as completely arbitrary. But it is then possible to put any two components of vector $\mathbf{r}_i$ equal to zero, for example y_i and z_i, and cancel out the nonzero component δx_i, which yields

$$m_i \frac{d^2 x_i}{dt^2} + \frac{\partial U}{\partial x_i} + \alpha \frac{\partial F}{\partial x_i} = 0 \tag{2.7}$$

In the same way we obtain a similar equation for any component. In vector form the equation is written as follows:

$$m_i \frac{d^2\mathbf{r}_i}{dt^2} + \frac{\partial U}{\partial \mathbf{r}_i} + \alpha \frac{\partial F}{\partial \mathbf{r}_i} = 0 \tag{2.8}$$

the subscript i numbering all the mass points of the mechanical system. Together with Eq. (2.2), Eqs. (2.8) make possible the determination of all $\mathbf{r}_i$'s (as functions of time) and the parameter α. Note that from (2.3) the products $-\alpha(\partial F/\partial \mathbf{r}_i)$ are in fact the reactions of the constraints.

We shall now formulate *Hamilton's principle*. For this transform the first term in Eq. (2.5) by parts:

$$m_i \frac{d^2\mathbf{r}_i}{dt^2} \delta\mathbf{r}_i = \frac{d}{dt} \left(m_i \frac{d\mathbf{r}_i}{dt} \delta\mathbf{r}_i \right) - m_i \frac{d\mathbf{r}_i}{dt} \frac{d}{dt} \delta\mathbf{r}_i$$

(we restrict ourselves for the time being to one term). Note that $\delta\mathbf{r}_i$ denotes the difference between two radius vectors taken at the same time. The derivative of a difference equals the difference between the derivatives, so that

$$\frac{d}{dt} \delta\mathbf{r}_i = \delta \frac{d\mathbf{r}_i}{dt}$$

Taking advantage of the fact that the symbol δ refers to an infinitesimal difference, we rewrite the equation as follows:

$$m_i \frac{d^2\mathbf{r}_i}{dt^2} \delta\mathbf{r}_i = \frac{d}{dt} \left(m_i \frac{d\mathbf{r}_i}{dt} \delta\mathbf{r}_i \right) - \delta \frac{m_i}{2} \left(\frac{d\mathbf{r}_i}{dt} \right)^2$$

We sum the obtained expression over the mass points, that is, over i. Thanks to the smallness of $\delta\mathbf{r}_i$, which can be regarded here as

the differential of the coordinate, the sum is

$$\sum_i \frac{\partial U}{\partial \mathbf{r}_i} \, \delta\mathbf{r}_i = \delta U$$

Collecting terms in Eq. (2.5), transformed in the way described, and taking the symbol δ outside the summation sign, we obtain

$$\frac{d}{dt} \left(\sum_i m_i \frac{d\mathbf{r}_i}{dt} \, \delta\mathbf{r}_i \right) - \delta \left[\sum_i \frac{m_i}{2} \left(\frac{d\mathbf{r}_i}{dt} \right)^2 - U \right] = 0 \qquad (2.9)$$

We assume now that the system is displaced according to the laws of mechanics from some given initial position which it occupied at time $t = t_0$ to another, also given, position at time t_1. Since the two positions are given, all $\delta\mathbf{r}_i$'s must vanish at t_0 and t_1:

$$(\delta\mathbf{r}_i)_{t=t_0} = (\delta\mathbf{r}_i)_{t=t_1} = 0 \qquad (2.10)$$

Integrate the expression on the left-hand side of (2.9) from t_0 to t_1. The total derivative with respect to time in this case reduces to the difference between the values of the differentiated quantity at the limits:

$$\sum_i m_i \left[\left(\frac{d\mathbf{r}_i}{dt} \, \delta\mathbf{r}_i \right)_{t=t_1} - \left(\frac{d\mathbf{r}_i}{dt} \, \delta\mathbf{r}_i \right)_{t=t_0} \right]$$

$$- \int_{t_0}^{t_1} \delta \left[\sum_i \frac{m_i}{2} \left(\frac{d\mathbf{r}_i}{dt} \right)^2 - U \right] dt = 0 \qquad (2.11)$$

But as we have pointed out, at the limits $\delta\mathbf{r}_i$ vanishes. Besides, the symbol δ, which denotes the difference between function values at the same instant, can be interchanged with the time integral for the very same reason that it is interchangeable with the time derivative. Denoting the integral itself by S, we arrive at the following equation:

$$\delta S = \delta \int_{t_0}^{t_1} \left[\sum_i \frac{m_i}{2} \left(\frac{d\mathbf{r}_i}{dt} \right)^2 - U \right] dt = 0 \qquad (2.12)$$

Since we used Eqs. (1.1) to derive Eq. (2.12), the integral in the expression for S is taken along the actual path. The symbol δ in front of the integral sign indicates that another integral was simultaneously evaluated along an infinitesimally close path spaced $\delta\mathbf{r}_i$ apart from the actual path for the ith particle. Such a close-lying path is said to be *varied*, and the symbol δ is the *variation* of the given quantity.

A variation has an entirely different meaning than a differential. The latter refers to the change in a quantity along the path of a

moving system, whereas the former corresponds to a transfer from one path to another, lying close to the initial one and compatible with the constraint imposed upon the system. A differential is determined from the equations of motion, a variation is subject only to the constraints and is otherwise arbitrary.

Equation (2.12) shows that the integral S taken along the actual path of the system possesses an extremum, since it does not change in passing over to a close path. Accordingly, near the extremum a function does not change its value when its argument is changed.

Instead of Eqs. (1.1) we can proceed from Eq. (2.12) as a fundamental proposition of mechanics. Such an approach may seem contrived. Actually, as we shall soon see, it opens the way to very broad generalizations. Besides, the equations of motion derived from condition (2.12) as a basic principle of mechanics can be much more suitable in various applications than the initial set of Eqs. (1.1). The quantity S is called the *action of a mechanical system*, and the assertion that S has an extreme value along the actual path is known as *Hamilton's principle*. In some cases the principle can be formulated in simpler terms, when it is known as the *principle of least action* (see Exercise at the end of Section 21).

Degrees of Freedom of a Mechanical System. In order to go over from rectangular coordinates to another coordinate system that is more convenient for solving certain mechanical problems, we must first formulate some essential general definitions.

Any independent parameter that defines the position of a mechanical system in space is known as its *degree of freedom*. The number of such independent parameters is called the *number of degrees of freedom* of the system.

The position of a single mass point in space is given by specifying three independent parameters (its coordinates) measured relative to a certain frame of reference. The position of N material points not joined by ideal constraints is defined by $3N$ independent parameters.

But if the configuration of the points is in some way secured, the number of degrees of freedom may be less than $3N$. For example, if two particles are joined by an ideal stationary constraint, their six Cartesian coordinates $(x_1, y_1, z_1, x_2, y_2, z_2)$ are subject to the condition

$$(x_1 - x_2)^2 + (y_1 - y_2)^2 + (z_1 - z_2)^2 = R_{12}^2$$

where R_{12} is the given distance between the points. Consequently not all the Cartesian coordinates are independent parameters: only five of these six quantities are independent. In other words, a system of two mass points at a constant distance from one another has five degrees of freedom.

If we consider three mass points rigidly joined by a triangle, the coordinates of the third point must satisfy two equations analogous to the one written above, with the quantities R_{13}^2 and R_{23}^2 in the right-hand sides. Thus the nine coordinates of the apexes of a rigid triangle are subject to three equations, and only six parameters are independent. The triangle has six degrees of freedom.

The position of a solid body in space is fully defined by three noncolinear points. Such three points, as was just shown, are given by six parameters. Hence, an arbitrary solid body has six degrees of freedom, provided only motions in which the body does not undergo deformation are considered (for instance, the rotation of a top).

Generalized Coordinates. The example of points joined by constraints shows that it is not always convenient to describe the position of a system in Cartesian coordinates, as this requires the writing of supplementary conditions due to the constraints. The choice of parameters needed to define the configuration of all the points of a mechanical system must be based primarily on considerations of expediency. Thus, if the forces depend only on the distances between the points, it is reasonable to introduce those distances into the dynamical equations explicitly and not in terms of Cartesian coordinates.

A mechanical system can be described by coordinates whose number is equal to the number of degrees of freedom of the system. These coordinates may sometimes coincide with the Cartesian coordinates of some of the particles. For example, in a system of two rigidly connected points, these coordinates can be chosen in the following way: the position of one of the points is given in Cartesian coordinates, after which the other point will always be situated on a sphere whose centre is the first point. The position of the second point on the sphere may be given by its longitude and latitude. Together with the three Cartesian coordinates of the first point, the latitude and longitude of the second point completely define the position of the system in space.

For three rigidly connected points, it is necessary, in accordance with the method just described, to specify the position of one side of the triangle and the angle of rotation of the third vertex about that side.

The independent parameters which define the position of a mechanical system in space are called its *generalized coordinates*. We will represent them by the symbols q_α, where the subscript α signifies the number of degrees of freedom.

The Lagrange Equations. Since generalized coordinates are independent, there is no need to impose constraint conditions upon them. This is one of their advantages over Cartesian coordinates

in solving dynamical problems. Another advantage appears when the generalized coordinates correspond to the symmetry properties of the system concerned. That is the case of spherical coordinates in the motion of a particle in a central force field. We shall now show how to write the equations of motion in generalized coordinates.

We could go straight over to them in Eqs. (1.1), but that is a cumbersome and not easily visualized procedure. It is much more convenient to proceed from Hamilton's principle (2.12). Since the generalized coordinates of a system fully define its spatial configuration, they can be used to express the Cartesian coordinates of its points. Let the transformation from Cartesian to generalized coordinates proceed according to the formulas

$$x_i = x_i (\ldots q_\alpha \ldots) \tag{2.13}$$

Differentiating, we obtain the expression for the Cartesian velocity components in terms of the derivatives dq_α/dt, called the *generalized velocities*. Instead of dq_α/dt we write, more briefly, $\dot{q}_\alpha$. We then obtain

$$\frac{dx_i}{dt} = \sum_\alpha \frac{\partial x_i}{\partial q_\alpha} \dot{q}_\alpha \tag{2.14a}$$

The summation is over all values of α, that is, over all the degrees of freedom of the system.

It will be readily observed that the index with respect to which the summation is performed is involved twice in the right-hand side of equation (2.14a): in the partial derivative and in the generalized velocity, the two quantities being multiplied. In such cases we shall not in future write the summation sign, assuming that the involvement of an index in a product twice signifies that the summation has been carried out. Such notation is not only space-saving but, given some practice, more easily visualized, since the formulas are not cluttered with summation symbols. The index with respect to which the summation is performed is called a *dummy index*. It can be redesignated on one side of the equation without touching the other. For example

$$\frac{dx_i}{dt} = \frac{\partial x_i}{\partial q_\alpha} \dot{q}_\alpha \equiv \frac{\partial x_i}{\partial q_\beta} \dot{q}_\beta \tag{2.14b}$$

The thing is that both α and β assume the same set of values, and it therefore does not matter what letter we write.

Let us now express x_i and dx_i/dt, which are involved in the integrand of S, in terms of the generalized coordinates and velocities according to formulas (2.13) and (2.14b). This expression (in any coordinate system) is known as the *Lagrange function* of a mechanical

system and is denoted L. Thus, let there be given the action of a system, S, such that $L = L\ (...\ q_\alpha\ ...\ \dot{q}_\alpha\ ...)$. Then

$$S = \int_{t_0}^{t_1} L\ (q_\alpha,\ \dot{q}_\alpha)\ dt \tag{2.15}$$

For the actual path of the system it possesses an extremum; this property cannot depend on the choice of the coordinate system since it expresses a known physical law. We now vary S, but in generalized, not Cartesian, coordinates. Since the former are independent, this is also true of their variations. We have

$$\delta S = \delta \int_{t_0}^{t_1} L\ dt = \int_{t_0}^{t_1} \delta L\ dt = \int_{t_0}^{t_1} \left(\frac{\partial L}{\partial q_\alpha} \delta q_\alpha + \frac{\partial L}{\partial \dot{q}_\alpha} \delta \dot{q}_\alpha \right) dt = 0$$
$$\tag{2.16}$$

Making use of the fact that the variation and differentiation symbols are commutative, we write

$$\frac{\partial L}{\partial \dot{q}_\alpha} \delta \dot{q}_\alpha \equiv \frac{\partial L}{\partial \dot{q}_\alpha}\ \delta \frac{dq_\alpha}{dt} = \frac{\partial L}{\partial \dot{q}_\alpha}\ \frac{d}{dt} \delta q_\alpha$$
$$= \frac{d}{dt} \left(\frac{\partial L}{\partial \dot{q}_\alpha} \delta q_\alpha \right) - \frac{d}{dt} \left(\frac{\partial L}{\partial \dot{q}_\alpha} \right) \delta q_\alpha \tag{2.17}$$

We integrate the total derivative with respect to time and substitute the limits. But at the limits the variations of the coordinates, as before, vanish, so that the following equation remains:

$$\delta S = \int_{t_0}^{t_1} \left[\frac{\partial L}{\partial q_\alpha} - \frac{d}{dt} \left(\frac{\partial L}{\partial \dot{q}_\alpha} \right) \right] \delta q_\alpha\ dt = 0 \tag{2.18}$$

Variations are mutually independent and arbitrary. We first put all δq_α's with the exception of δq_1 equal to zero. Then Eq. (2.18) retains only the first term in the summation with respect to α:

$$\delta S = \int_{t_0}^{t_1} \left[\frac{\partial L}{\partial q_1} - \frac{d}{dt} \left(\frac{\partial L}{\partial \dot{q}_1} \right) \right] \delta q_1\ dt = 0 \tag{2.19}$$

We now take advantage of the arbitrariness of the variation δq_1. Suppose that the quantity in brackets, by which δq_1 is multiplied, in some way changes its sign and absolute value but does not vanish over the integration interval. We now select δq_1 such that it is everywhere of the same sign as the expression in brackets. Then the integrand is positive, so that δS cannot vanish. Hence, for Hamilton's principle to hold the expression multiplied by δq_1 must of

necessity vanish. Using the same reasoning for an arbitrary δq_α, we find that

$$\frac{d}{dt}\frac{\partial L}{\partial \dot{q}_\alpha} - \frac{\partial L}{\partial q_\alpha} = 0 \tag{2.20}$$

This is the equation of motion written in forms of generalized coordinates. The set of Eqs. (2.20) is known as the *Lagrange equations of motion*.

If the number of degrees of freedom is n, then to integrate the Lagrange equations, which are of the second order with respect to time, $2n$ initial conditions must be given: n generalized coordinates and n generalized velocities for time t_0. Each generalized coordinate will then be a function of time, the initial velocities, and the initial coordinates:

$$q_\alpha = q_\alpha\,(t;\ q_{01},\ \ldots,\ q_{0n};\ \ \dot{q}_{01},\ \ldots,\ \dot{q}_{0n}) \tag{2.21}$$

Differentiating these equations with respect to time, we obtain the generalized velocities as functions of the same quantities:

$$\dot{q}_\alpha = \dot{q}_\alpha\,(t;\ q_{01},\ \ldots,\ q_{0n};\ \ \dot{q}_{01},\ \ldots,\ \dot{q}_{0n}) \tag{2.22}$$

Elimination of all the initial values of the coordinates and velocities, that is, solution of Eqs. (2.21) and (2.22) with respect to the initial coordinates and velocities, yields $2n$ equations of the form

$$q_{0\alpha}(t;\ q_1,\ \ldots,\ q_n;\ \ \dot{q}_1,\ \ldots,\ \dot{q}_n) = q_{0\alpha} = \text{constant} \tag{2.23}$$

Functions of the coordinates and velocities of a system that remain constant throughout the motion are known as the *integrals of the motion*. In the right-hand side they can have any constant coordinates and velocities, which need not necessarily be the initial ones. Determination of the integrals of the motion is one of the problems of mechanics.

The Determinancy of the Lagrange Function. As is apparent from its definition, the Lagrange function (or simply Lagrangian), contains two terms:

$$L = \sum_i \frac{m_i}{2}\left(\frac{d\mathbf{r}_i}{dt}\right)^2 - U \tag{2.24a}$$

The first term, which is quadratically dependent upon the velocities, is called the *kinetic energy* of the system; the second, which describes the interaction between particles, is called the *potential energy*. The meaning of both will be made clearer in Section 4.

The Lagrange equation (2.20) involves not the function L itself but its derivatives. This gives rise to the question of the determinancy

of L, that is, of possible supplementary terms not affecting the equations of motion. It is obvious, for example, that in any case an additional constant term does not affect the equations of motion; also, a total time derivative of any function of all q_α's and t's can be added to the Lagrangian without in any way affecting the system of equations (2.20).

This is easily verified by direct substitution as well as in the following simple way. The term $df\,(q_\alpha,\,t)/dt$, which has the form of a total derivative, can be integrated; as a result it is added to the action in the form of a difference between the function values at the limits:

$$S = \int_{t_0}^{t_1} \left(L + \frac{df}{dt} \right) dt = \int_{t_0}^{t_1} L\,dt + f\,(q_\alpha,\,t) \Big|_{t_0}^{t_1} \tag{2.25}$$

But since the variations of the coordinates vanish at the limits, these values remain constant in the variation of q_α. Hence the derivative of a function of coordinates and time is not involved in the variation of the action and does not affect the equations of motion. This property can be used to determine the form of L, if it is not given in advance as (2.24a), on the basis of Hamilton's principle and certain other general propositions of mechanics.

The Principle of Relativity. The concept of an inertial frame of reference was defined above as a frame in which all the accelerations of particles are due solely to interactions between them. Suppose we have such a frame. Then all other inertial reference frames must be moving uniformly in a straight line relative to it. Otherwise bodies moving relative to the initial frame with a velocity constant in magnitude and direction would be found to be moving with an acceleration relative to another reference frame. But in that case the latter would not by definition be inertial.

Thus, all inertial reference frames are in rectilinear uniform motion relative to one another. Any one of them can be legitimately assumed at rest, and all the others moving. The equations of motion of a mechanical system have the same form in any inertial reference frame. A common example is that of a passenger in a train travelling at a uniform velocity: he sees all physical phenomena in the coach exactly the same as if the train were at rest. It would be better to say that this is not an example demonstrating the equivalence of two inertial frames of reference but experimental proof of the fundamental mechanical principle known as the *principle of relativity.* As applied to Newtonian mechanics, which reflects simple facts known to us from everyday life, the principle seems self-evident. But when it was applied to the theory of electromagnetism, it led to a fundamental revision of physical concepts (see Part II).

The Symmetry of the Laws of Motion. The property whereby an equation expressing a known physical relationship retains its form in a transformation is known as *symmetry* with respect to that transformation. The relativity principle declares that the equations of motion are symmetrical with respect to the substitution of one inertial frame for another. Experience shows that the laws of mechanics possess other forms of symmetry as well.

In a mechanical system that is sufficiently distant from other bodies motion is always the same wherever the system is located. This means the following. Let there be two identical mechanical systems with identical initial conditions of motion, both of which are very far away from any other bodies capable of affecting them. In that case, if they are taken in the same reference frame, motion in them occurs in strictly the same way. In other words, the motion is not affected by the transfer of all the moving bodies over the same distance, along parallel lines, at the same time. This assertion is, of course, based on the vast experience accumulated by mechanics in the whole course of its development. More briefly the property is known as the *homogeneity of space.*

Two equivalent mechanical systems like the ones described here can be taken not only displaced relative to one another but also turned through any angle. Again, if the two systems are sufficiently far away from all bodies capable of affecting them, motion in them takes place in the same way. In other words all directions in space are equivalent. This property of space is known as the *isotropy of space*. Like homogeneity, the isotropy of space also follows from the sum-total of experience. Homogeneity and isotropy are an expression of specific properties of the laws of motion: their symmetry with respect to displacements and rotations. Mathematically, displacements and rotations in space are represented by corresponding transformations of the coordinate system.

There is one more type of symmetry of the laws of motion. They are homogeneous with respect to time transfer: the laws of motion do not change with time. If this property of the laws of motion of mechanical systems did not hold, it would be impossible to design any machine.

Determination of the Form of the Lagrangian. The laws of symmetry of motion listed above, that is, space and time homogeneity, space isotropy, the relativity principle, and Hamilton's principle, can be used to determine the form of the Lagrangian without preliminary reference to Eqs. (1.1).

Let us start with a free particle sufficiently far away from all other bodies (which is the definition of a free particle). By virtue of space homogeneity, its Lagrangian cannot be explicitly dependent on the coordinates, since otherwise at different spatial points the

particle would move according to different laws. For the same reason the Lagrangian does not explicitly involve time, and this refers not only to an individual free particle but to any assembly of particles not subject to external forces. Thus the Lagrangian of a free particle can depend only upon its velocity. But L is a scalar quantity. A scalar can be obtained from a vector in one of two ways: by taking the absolute value of the vector or by multiplying it scalarly by another vector. But there is no such preferred vector in isotropic space since all directions in it are equivalent. Thus the only possible form of the Lagrangian of a free particle is $L = L\,(|\,\dot{\mathbf{r}}\,|)$.

It remains to determine what the function is. According to the relativity principle the character of motion should not change in passing to another inertial reference frame. As was pointed out, the latter must be travelling rectilinearly and uniformly relative to the initial one. If its velocity is $\mathbf{V}$, the particle under consideration moves relative to it with the velocity $\dot{\mathbf{r}} + \mathbf{V}$. We have made use of the simple law of velocity composition which, as will be shown in Part II, holds only when both $|\,\dot{\mathbf{r}}\,|$ and $|\,\mathbf{V}\,|$ are substantially below the speed of light. Thus in the new inertial frame the Lagrangian is $L = L\,(|\,\dot{\mathbf{r}} + \mathbf{V}\,|)$. For the law of motion to remain the same the difference between the two expressions must be equal to the total derivative of a certain function of the coordinates and time. It is immediately apparent that for a free particle this leaves only one possibility:

$$L = \frac{m}{2}\,|\,\dot{\mathbf{r}}\,|^2 \qquad\qquad (2.26)$$

where m is a constant quantity.

Indeed, we then obtain

$$\frac{m\,|\,(\dot{\mathbf{r}} + \mathbf{V})\,|^2}{2} - \frac{m\,|\,\dot{\mathbf{r}}\,|^2}{2} = m\mathbf{V}\dot{\mathbf{r}} + \frac{m\,|\,\mathbf{V}\,|^2}{2}$$

$$= \frac{d}{dt}\left(m\mathbf{r}\mathbf{V} + \frac{m\,|\,\mathbf{V}\,|^2\,t}{2}\right)$$

What is the sign of m? Let us determine it. But first we must somewhat refine Hamilton's principle by requiring that along short paths the action be not simply extremal but minimal. Then the sign of m is positive. At negative m the action could decrease limitlessly with the increase of $|\,\dot{\mathbf{r}}\,|$. We have thus finally determined the first term in the Lagrangian for a free particle.

If we now take a system of interacting particles, to describe their interaction we must introduce an additional term into the Lagrangian.

We assumed that the action of the particles on one another depends only on their position at a given time. It is, however, significant that it is determined only by their relative position, that is, it depends only on the separation $r_i - r_h$ and not on each vector separately. Only the differences between vectors remain constant in transfers of a coordinate system. In addition, only the differences $r_i - r_h$ satisfy the relativity principle: the products Vt, which are added to each radius vector r_i, r_h in going over to another inertial frame, cancel out.

Since the Lagrangian is a scalar quantity, it can depend either on the absolute values of the differences $| r_i - r_h |$ or on scalar products of the type $(r_i - r_h)(r_l - r_m)$. But the latter case is not encountered in practice and need not be considered. Hence the Lagrangian of a system of material points not interacting with other bodies is

$$L = \sum_i \frac{m_i}{2} (\dot{r}_i)^2 - U \left(\ldots | r_i - r_h | \ldots \right) \qquad (2.24b)$$

We have not restricted ourselves to developing the Lagrange equations from Eqs. (1.1) and we have performed all of the complex reasoning needed to deduce formula (2.24b) because in this way it is easier to arrive at the necessary generalizations required by Einstein's relativity principle and electromagnetic field theory.

The special significance of Hamilton's principle in mechanics consists in that it makes it possible to express all symmetry properties of mechanical systems in the most clear and concise form. Although they can be derived from the differential equations of motion as well, the integral principle expresses them much more distinctly. Since the symmetry of the conditions of motion is a generalization of certain experimentally established laws, Hamilton's principle provides the most convenient means of formulating all the general laws of mechanics. It should, of course, be borne in mind that this formulation is a reflection of a tendency to seek the most concise and convenient notation, not of any natural "striving" for minimum action.

Symmetry properties substantially restrict the possible behaviour of mechanical systems. As will be shown later in Section 4, different types of symmetry are associated with certain quantities (dependent upon dynamic variables) whose values, determined at the initial instant of time, are conserved. This substantially restricts the domain of variables in the problems considered. In a number of important cases these quantities are best found with the help of the variation principle, according to the symmetry properties inherent in it.

Hamilton's principle, with due account of symmetry requirements, can be used to determine the form of the Lagrangian, and thereby the form of the equations of motion. In this sense it possesses great heuristic force, that is, makes it possible to find unknown quantities from general considerations.

Finally, the variation principle is extremely convenient in solving specific problems of mechanics with the help of the Lagrange equations obtained by variation.

3

EXAMPLES OF CONSTRUCTING THE LAGRANGE EQUATIONS

The Rules for Constructing the Lagrange Equations. Let us sum up the sequence of operations for developing the Lagrange equations for a specific mechanical system:

(i) Express the Cartesian coordinates in terms of the generalized coordinates:

$$x_i = x_i \, (q_1, \ \ldots, \ q_\alpha, \ \ldots, \ q_n)$$

(ii) By differentiating these equalities obtain the Cartesian velocity components expressed in terms of the generalized coordinates and generalized velocities (bearing in mind the summation rule for the subscript α):

$$\dot{x}_i = \frac{\partial x_i}{\partial q_\alpha} \dot{q}_\alpha$$

(iii) Substitute generalized coordinates for the Cartesian coordinates involved in the potential energy formula:

$$U \, (\ldots \mid x_i - x_k \mid \ldots) = U \, (q_1, \ \ldots, \ q_\alpha, \ \ldots, \ q_n)$$

(iv) Substitute the generalized velocities for the velocities involved in the kinetic energy formula, so that in the most general case the kinetic energy becomes dependent not only on the generalized velocities but on the generalized coordinates as well:

$$T \equiv \sum_i \frac{m_i}{2} (\dot{\mathbf{r}}_i)^2 = \frac{1}{2} \sum_i m_i \frac{\partial x_i}{\partial q_\alpha} \dot{q}_\alpha \frac{\partial x_i}{\partial q_\beta} \dot{q}_\beta \equiv \frac{1}{2} T_{\alpha\beta} \dot{q}_\alpha \dot{q}_\beta$$

Since the Cartesian velocity components are homogeneous linear functions of the generalized velocities $\dot{q}_\alpha$, the kinetic energy is a quadratic homogeneous function of $\dot{q}_\alpha$ and $\dot{q}_\beta$.

(v) Compute the partial derivatives $\partial L/\partial \dot{q}_\alpha$ and $\partial L/\partial \dot{q}_\beta$, assuming $\dot{q}_\alpha$ and $\dot{q}_\beta$ to be independent variables.

(vi) Substitute the derivatives for all degrees of freedom into Eqs. (2.20).

Let us now examine several examples of developing the Lagrange equations.

Central Forces. This is the name given to forces directed along lines joining the mass points and dependent only on the distances between the particles. Actually we already considered such forces when we assumed that potential energy depends only on the distances between points. Then the force with which the kth point acts on the ith is

$$\mathbf{F}_{ik} = -\frac{\partial}{\partial \mathbf{r}_{ik}} U \left(\ldots |\mathbf{r}_i - \mathbf{r}_k| \ldots \right) = -\frac{\partial U}{\partial |\mathbf{r}_{ik}|} \frac{\partial |\mathbf{r}_{ik}|}{\partial \mathbf{r}_{ik}} \quad (3.1)$$

The derivative of the scalar quantity $|\mathbf{r}_i - \mathbf{r}_k| \equiv |\mathbf{r}_{ik}|$ with respect to vector $\mathbf{r}_{ik}$ is a vector with components

$$\frac{\partial |\mathbf{r}_{ik}|}{\partial x_{ik}}, \qquad \frac{\partial |\mathbf{r}_{ik}|}{\partial y_{ik}}, \qquad \frac{\partial |\mathbf{r}_{ik}|}{\partial z_{ik}}$$

For an example let us compute the component along the x axis:

$$\frac{\partial}{\partial x_{ik}} \left(x_{ik}^2 + y_{ik}^2 + z_{ik}^2 \right)^{1/2} = \frac{x_{ik}}{|\mathbf{r}_{ik}|} \quad (3.2)$$

But the ratio $x_{ik}/|\mathbf{r}_{ik}|$ is the component along the x axis of a unit vector directed from the ith particle to the kth, which implies that the force represented by expression (3.1) is a central force.

If one of the bodies of a system is much more massive than all the others (like the sun in our solar system) it can, to a certain approximation, be assumed at rest, that is, the actions of the other bodies on it can be neglected. In such cases the body is said to produce a *field* in which the mass points are moving.

Suppose we have a gravitational field, so that the central body attracts the others with a force inversely proportional to the distance from it. Neglecting the action of the bodies on each other, it is easy to find the expression for the potential energy of a particle in the field of the central body. The force with which it acts on the particle is

$$\mathbf{F} = -\frac{a}{r^2} \frac{\mathbf{r}}{r} \quad (3.3)$$

where a is a constant factor.

Comparing (3.3) with Eqs. (3.1) and (3.2), we conclude that the derivative of the potential energy with respect to the distance is

$$\frac{\partial U}{\partial r} = \frac{a}{r^2} \quad (3.4)$$

Integrating this equation, we find

$$U = \text{constant} - \frac{a}{r} \qquad (3.5)$$

The constant here is the value of the potential energy at an infinite distance from the attracting body. In the case of forces decreasing with distance this constant is usually taken to be zero, but this is a pure formality, since when a force is computed by differentiating the potential energy the constant is eliminated anyway.
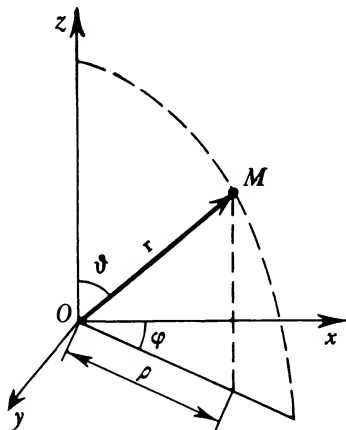

Figure 1

The convention concerning the choice of the constant in the potential energy expression is called gauging. In this case, at infinity from the attracting body U is gauged to zero:

$$U = -\frac{a}{r} \qquad (3.6)$$

An expression analogous to (3.5) is also obtained for two points carrying an electric charge, so that the Coulomb law is, in form, like Newton's law of gravitation. However, since it corresponds to two signs of charges it can denote either attraction or repulsion. Correspondingly, the potential energy of the Coulomb forces has both signs: positive for like charges and negative for opposite.

Spherical Coordinates. Formula (3.6) suggests that in this instance it is best to choose r as the generalized coordinate. In other words, we must transform from Cartesian to spherical coordinates. The relationship between Cartesian and spherical coordinates is shown in Figure 1. The z axis is called the polar axis of the spherical coor-

dinate system. The angle ϑ between the radius vector and the polar axis is called the polar angle; it is complementary (to 90°) to the "latitude". Finally, the angle φ is analogous to the "longitude" and is called the azimuth. It measures the dihedral angle between the plane zOx and the plane passing through the polar axis and the given point M.

Let us find the formulas for the transformation from Cartesian to spherical coordinates. From Figure 1 it is clear that

$$z = r \cos \vartheta \tag{3.7}$$

The projection ρ of the radius vector onto the plane xOy is

$$\rho = r \sin \vartheta \tag{3.8a}$$

Whence

$$x = \rho \cos \varphi = r \sin \vartheta \cos \varphi \tag{3.8b}$$

$$y = \rho \sin \varphi = r \sin \vartheta \sin \varphi \tag{3.8c}$$

We shall now find the expression for the kinetic energy in spherical coordinates. This can be done either by direct calculation according to the method indicated at the beginning of this section or by a geometrical construction. Although the latter is simpler, let us first follow the computation procedure in order to illustrate the general method. We have

$$\dot{z} = \dot{r} \cos \vartheta - r \sin \vartheta \, \dot{\vartheta}$$

$$\dot{x} = \dot{r} \sin \vartheta \cos \varphi + r \cos \vartheta \cos \varphi \, \dot{\vartheta} - r \sin \vartheta \sin \varphi \, \dot{\varphi}$$

$$\dot{y} = \dot{r} \sin \vartheta \sin \varphi + r \cos \vartheta \sin \varphi \, \dot{\vartheta} + r \sin \vartheta \cos \varphi \, \dot{\varphi}$$

Squaring these equations and adding, we obtain, after very simple manipulations,

$$T = \frac{1}{2} m (\dot{x}^2 + \dot{y}^2 + \dot{z}^2) = \frac{m}{2} (\dot{r}^2 + r^2\dot{\vartheta}^2 + r^2 \sin^2 \vartheta \, \dot{\varphi}^2) \tag{3.9}$$

The same is clear from the construction shown in Figure 2. An arbitrary displacement of the point can be resolved into three mutually perpendicular displacements: dr, $r\,d\vartheta$ and $\rho\,d\varphi = r \sin \vartheta \, d\varphi$. Whence

$$dl^2 = dr^2 + r^2 \, d\vartheta^2 + r^2 \sin^2 \vartheta \, d\varphi^2 \tag{3.10}$$

Since the square of the velocity $v^2 = (dl^2/dt)^2$, (3.9) is obtained from (3.10) simply by dividing by $(dt)^2$ and multiplying by $m/2$.

Hence, in spherical coordinates, the Lagrangian is expressed as

$$L = \frac{m}{2} (\dot{r}^2 + r^2 \sin^2 \vartheta \, \dot{\varphi}^2 + r^2 \dot{\vartheta}^2) - U(r) \tag{3.11}$$

Now in order to write down the Lagrange equations it is sufficient to calculate the partial derivatives:

$$\frac{\partial L}{\partial \dot{r}} = m\dot{r}, \qquad \frac{\partial L}{\partial \dot{\vartheta}} = mr^2\dot{\vartheta}, \qquad \frac{\partial L}{\partial \dot{\varphi}} = mr^2 \sin^2 \vartheta \, \dot{\varphi}$$

$$\frac{\partial L}{\partial r} = mr \sin^2 \vartheta \, \dot{\varphi}^2 + mr\dot{\vartheta}^2 - \frac{\partial U}{\partial r}, \qquad \frac{\partial L}{\partial \varphi} = 0$$

$$\frac{\partial L}{\partial \vartheta} = mr^2 \sin \vartheta \cos \vartheta \, \dot{\varphi}^2$$

These derivatives must be substituted into (2.20), which, however, we shall not now do since the motion we are considering actually reduces to the plane case (see beginning of Section 5).
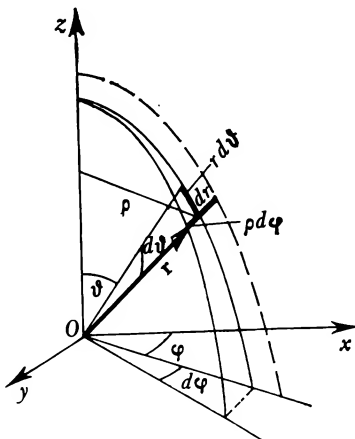

Figure 2

Two-particle Systems. So far we have considered the centre of attraction as stationary, which corresponds to the assumption of an infinitely large mass. But it may happen that both masses are similar or equal to each other (a binary star, a neutron-proton system, and the like). We shall show that the problem of the motion of two bodies interacting only with one another can always be easily reduced to a problem of the motion of a single body.

Let the mass of the first particle be m_1 and of the second m_2. We call the radius vectors of these particles, drawn from an arbitrary origin, $\mathbf{r}_1$ and $\mathbf{r}_2$, respectively. The components of $\mathbf{r}_1$ are x_1, y_1, z_1; the components of $\mathbf{r}_2$ are x_2, y_2, z_2. We now define the radius vector of the centre of mass of these particles, $\mathbf{R}$, by the following formula:

$$\mathbf{R} = \frac{m_1\mathbf{r}_1 + m_2\mathbf{r}_2}{m_1 + m_2} \qquad (3.12)$$

Sometimes the terms "centre of inertia" or "centre of gravity" are used. However, the centre of gravity can be determined only in a uniform field of gravitational forces.

In addition, let us introduce the radius vector of the relative position of the particles

$$\mathbf{r} = \mathbf{r}_1 - \mathbf{r}_2 \tag{3.13}$$

Let us express their kinetic energy in terms of $\dot{\mathbf{R}}$ and $\dot{\mathbf{r}}$. Eliminating $\mathbf{r}_1$ and $\mathbf{r}_2$ from (3.12) and (3.13), and differentiating them with respect to time, we obtain

$$\dot{\mathbf{r}}_1 = \dot{\mathbf{R}} + \frac{m_2}{m_1 + m_2}\dot{\mathbf{r}}, \qquad \dot{\mathbf{r}}_2 = \dot{\mathbf{R}} - \frac{m_1}{m_1 + m_2}\dot{\mathbf{r}} \tag{3.14}$$

The kinetic energy is equal to

$$T = \frac{m_1}{2}\dot{\mathbf{r}}_1^2 + \frac{m_2}{2}\dot{\mathbf{r}}_2^2 \tag{3.15}$$

Substituting (3.14) into it, we obtain, after a simple rearrangement,

$$T = \frac{m_1 + m_2}{2}\dot{\mathbf{R}}^2 + \frac{m_1 m_2}{2(m_1 + m_2)}\dot{\mathbf{r}}^2 \tag{3.16}$$

The cross term involving $\dot{\mathbf{R}}\dot{\mathbf{r}}$ has been eliminated, which is the purpose of the transformation.

Since by definition there are no external forces acting on the mass points, the potential energy can depend only on the distance between them, r: $U = U(r)$. Thus the Lagrangian is

$$L = \frac{m_1 + m_2}{2}\dot{\mathbf{R}}^2 + \frac{m_1 m_2}{2(m_1 + m_2)}\dot{\mathbf{r}}^2 - U(r) \tag{3.17}$$

Let us now write the Lagrange equations for the centre-of-mass coordinates. Differentiation of Eq. (3.17) yields

$$\frac{\partial L}{\partial \dot{\mathbf{R}}} = (m_1 + m_2)\dot{\mathbf{R}}, \qquad \frac{\partial L}{\partial \mathbf{R}} = 0$$

Hence, in accordance with (2.20), we have $\ddot{\mathbf{R}} = 0$, or in Cartesian coordinates

$$\ddot{X} = 0, \qquad \ddot{Y} = 0, \qquad \ddot{Z} = 0$$

These equations can be easily integrated. Whence

$$X = \dot{X}_0 t + X_0, \qquad Y = \dot{Y}_0 t + Y_0, \qquad Z = \dot{Z}_0 t + Z_0$$

where the subscript 0 corresponds to the values of the quantities at time zero. Combining the coordinate equations into one vector equation, we obtain

$$\mathbf{R} = \dot{\mathbf{R}}_0 t + \mathbf{R}_0 \tag{3.18}$$

Thus, the centre of mass moves uniformly in a straight line relative to the initial reference frame. But we assumed the frame to be inertial because the forces in it are due only to interactions between the mass points [the corresponding potential energy is $U(r)$]. Hence, a reference frame fixed with respect to the centre of mass of two mass points is also inertial. If we pass to it, there remains the relative motion of both mass points, which is described by the separation $\mathbf{r}$. In the centre-of-mass reference frame the Lagrangian has the form

$$L_{\text{c.m}} = \frac{m_1 m_2}{2(m_1 + m_2)} \dot{\mathbf{r}}^2 - U(r) \tag{3.19}$$

Here, obviously, $\dot{r}^2 = \dot{x}^2 + \dot{y}^2 + \dot{z}^2$. This Lagrangian involves only three coordinates, not six as in (3.17). Consequently the problem of the motion of two bodies of masses m_1 and m_2 reduces to the problem of the motion of one body of mass

$$m = \frac{m_1 m_2}{m_1 + m_2} \tag{3.20}$$

which is called the *reduced mass*.

Since the reference frame connected with the centre of mass is inertial, the motion of the centre of mass does not affect the relative motion of the mass points. In the next section we shall show that this assertion holds for any number of mass points not subject to external forces. It can simply be assumed that the centre of mass is at rest at the origin of the coordinate system, $\mathbf{R} = 0$.

If the relative motion of two mass points is described in spherical coordinates, the equations of motion have the same form as for one point moving relative to a fixed centre of attraction.

Assuming the centre of mass of the two points to be at rest at the origin of the coordinate system, we find the distances of both points from the origin:

$$r_1 = \frac{m_2 r}{m_1 + m_2}, \qquad r_2 = \frac{m_1 r}{m_1 + m_2}$$

Thus, if one of the masses is much smaller than the other ($m_2 \ll m_1$), then $r_1 \ll r_2$, that is, the centre of mass of the two points lies very close to the point of greater mass. This is the case for a system consisting of the earth and an artificial satellite and, to a smaller approximation, for the earth-moon system. The reduced mass can be written thus:

$$m = \frac{m_2}{1 + m_2/m_1} \tag{3.21}$$

From this it can be seen that it equals approximately the smaller mass. That is why the motion of a satellite can be described as if the earth were fixed and the mass of the satellite were independent of the mass of the earth.

Simple and Double Pendulums. In concluding this section we shall derive the Lagrangian for simple and double pendulums. We shall make use of them later.

The simple plane pendulum is a mass point m suspended on a flat hinge at a certain point by a weightless rod of length l. The hinge restricts the plane of oscillation of the pendulum (Figure 3). It is clear that such a pendulum has one degree of freedom. We take
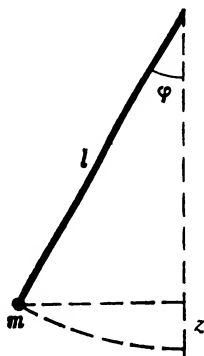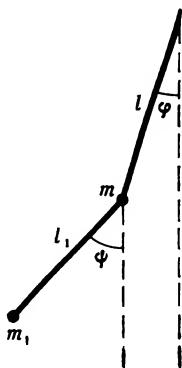

Figure 3   Figure 4

the angle of deflection of the pendulum from the vertical, φ, as a generalized coordinate. Obviously the velocity of the mass point is equal to $l\dot{\varphi}$, so that the kinetic energy is

$$T = \frac{m}{2}\, l^2\dot{\varphi}^2$$

Acting on the pendulum is the force of gravity $-mg$. Here, g is the acceleration of free fall; the minus sign takes account of the fact that the force is directed downwards. Hence the potential energy $U = mgz$, and $F = -\partial U/\partial z$, where z is expressed in terms of the angle as follows: $z = l\,(1 - \cos\varphi)$. Thus the Lagrangian of the pendulum is

$$L = \frac{m}{2}\, l^2\dot{\varphi}^2 - mgl\,(1 - \cos\varphi) \tag{3.22}$$

Note that mass m is involved as a common multiplier of both terms in expression (3.22) and therefore cancels out in the equations of motion. It follows that the law of oscillation of a pendulum does not depend on its mass. This, of course, is true if all types of friction can be neglected.

A double pendulum can be described in the following way: at point m there is another hinge from which another pendulum, which

is forced to oscillate in the same plane (Figure 4), is suspended. Let the mass and length of the second pendulum be m_1 and l_1, respectively, and its angle of deflection from the vertical, ψ. The coordinates of the second mass point are

$$x_1 = l \sin \varphi + l_1 \sin \psi$$

$$z_1 = l \, (1 - \cos \varphi) + l_1 \, (1 - \cos \psi)$$

Whence we obtain its velocity components:

$$\dot{x}_1 = l \cos \varphi \, \dot{\varphi} + l_1 \cos \psi \, \dot{\psi}$$

$$\dot{z}_1 = l \sin \varphi \, \dot{\varphi} + l_1 \sin \psi \, \dot{\psi}$$

Squaring and adding them, we express the kinetic energy of the second particle in terms of the generalized coordinates φ, ψ and the generalized velocities $\dot{\varphi}$, $\dot{\psi}$:

$$T_1 = \frac{m_1}{2} \, [l^2 \dot{\varphi}^2 + l_1^2 \dot{\psi}^2 + 2 l l_1 \cos (\varphi - \psi) \, \dot{\varphi} \dot{\psi}]$$

The potential energy of the second particle is determined in terms of z_1. Finally, we get an expression of the Lagrangian for a double pendulum in the following form:

$$L = \frac{m + m_1}{2} \, l^2 \dot{\varphi}^2 + \frac{m_1}{2} \, l_1^2 \dot{\psi}^2 + m_1 l l_1 \cos (\varphi - \psi) \, \dot{\varphi} \, \dot{\psi}$$

$$- (m + m_1) \, g l \, (1 - \cos \varphi) - m_1 g l_1 \, (1 - \cos \psi) \quad (3.23)$$

By using generalized coordinates we completely avoided the question of the forces of reaction appearing in the hinges.

EXERCISES

1. Write the Lagrange equation if the Lagrangian has the form

$$L = - (1 - \dot{q}^2)^{1/2}$$

We shall encounter Lagrangians of a similar type in Part II.

2. A point moves in a vertical plane along a given curve in a gravitational field. The equation of the curve in parametric form is $x = x(s)$, $z = z(s)$. Write the Lagrange equations.

Solution. The velocities are

$$\dot{x} = \frac{dx}{ds} \, \dot{s} \equiv x' \dot{s}, \qquad \dot{z} = \frac{dz}{ds} \, \dot{s} \equiv z' \dot{s}$$

The Lagrangian has the form

$$L = \frac{m}{2}(x'^2 + z'^2)\,\dot{s}^2 - mgz(s)$$

The Lagrange equation is

$$\frac{d}{dt}\,m\,[(x'^2 + z'^2)\,\dot{s}] - m\dot{s}^2\,(x'x'' + z'z'') + mgz' = 0$$

3. Write the Lagrange equations for an elastically suspended pendulum. The potential energy of the elastic force in the case of such a pendulum is calculated from the formula $U = (k/2)(l - l_0)^2$, where l_0 is the equilibrium length of the unextended rod, and k is a constant characterizing its elasticity. Use l and φ as the generalized coordinates.

4. Write the kinetic energy of a system of three mass points m_1, m_2, m_3 in terms of the kinetic energy of motion of the three points moving together with the centre of mass and the kinetic energy of relative motion.

Answer.

$$T = \frac{M\dot{\mathbf{R}}^2}{2} + \frac{m_2(m_1 + m_3)}{2M}\,\dot{\rho}_2^2 + \frac{m_3(m_1 + m_2)}{2M}\,\dot{\rho}_3^2 - \frac{m_2 m_3}{M}\,\dot{\rho}_2\dot{\rho}_3$$

where

$$M = m_1 + m_2 + m_3, \qquad \rho_2 = \mathbf{r}_1 - \mathbf{r}_2, \qquad \rho_3 = \mathbf{r}_1 - \mathbf{r}_3$$
$$\mathbf{R} = \frac{m_1\mathbf{r}_1 + m_2\mathbf{r}_2 + m_3\mathbf{r}_3}{M}$$

4

CONSERVATION LAWS

In Section 2 we gave a general definition of the integrals of the motion. Finding all the integrals of the motion of an arbitrary mechanical system is extremely difficult and rarely accomplished in analytical form. However, there are certain important integrals of the motion which can be written directly according to the form of the Lagrangian. These integrals will be examined in this section.

Energy. Let us use the Lagrangian to determine a quantity

$$E = \dot{q}_\alpha\,\frac{\partial L}{\partial \dot{q}_\alpha} - L \tag{4.1}$$

(the summation over α is from 1 to n). The quantity E is called the *total energy* of a system. Let us calculate its total derivative with

respect to time:

$$\frac{dE}{dt} = \ddot{q}_\alpha \frac{\partial L}{\partial \dot{q}_\alpha} + \dot{q}_\alpha \frac{d}{dt} \frac{\partial L}{\partial \dot{q}_\alpha} - \dot{q}_\alpha \frac{\partial L}{\partial q_\alpha} - \ddot{q}_\alpha \frac{\partial L}{\partial \dot{q}_\alpha} - \frac{\partial L}{\partial t}$$

The last three terms in the right-hand side are the derivative of the Lagrangian, dL/dt, which is dependent on q_α, $\dot{q}_\alpha$, and in some cases on time. From the Lagrange equations, the quantity $(d/dt)(\partial L/\partial \dot{q}_\alpha)$ can be expressed as $\partial L/\partial q_\alpha$. Thus

$$\frac{dE}{dt_c} = -\frac{\partial L}{\partial t} \tag{4.2}$$

Hence, if the Lagrangian is not explicitly dependent on time, the energy E is an integral of the motion, that is, it is conserved.

Let us now consider in what cases L does not depend on t. It was pointed out in Section 2 that if a given mechanical system is sufficiently far away from all other bodies, or *closed*, as they say, then time is not involved in the Lagrangian. This expresses the *homogeneity of time*. That means that in a closed system energy is conserved, or as we have agreed to say, it is an integral of the motion. If there are two noninteracting closed systems, the Lagrangian for them comprises the sum of the Lagrangians of each system separately. Accordingly, the total energy is, by definition (4.1), expressed as the sum of the energies of both systems. The energy of such systems is an additive integral of the motion.

Energy is conserved not only in isolated, closed systems. If there is a uniform external force field acting on a system, the Lagrangian of such a system does not involve time explicitly, so that, from (4.2), the energy is also an integral of the motion. When constraints are imposed upon a system, its Cartesian coordinates are expressed in terms of the generalized coordinates according to formulas that do not explicitly involve time. This case was examined in Section 2. The energy of the system here is also conserved.

Ideal constraints can be treated as a special case of a field of force, so that the action of constraints not involving time is analogous to a uniform field. Therefore energy is in this case conserved. But in a variable external field or with constraints explicitly involving time the energy of a system is not conserved: either work is done on the system or the system itself does work on some external object.

When forces of friction act within a closed system, the energy of macroscopic motion transforms into the energy of microscopic, molecular, motion. Together with this internal energy, the energy of a closed system is, of course, conserved, but the Lagrangian, which involves only the generalized coordinates of motion of the system as a whole, no longer provides a complete description of the system's motion. The mechanical energy of the macroscopic motion

alone, determined with the help of such a Lagrangian, is no longer conserved. Mechanical energy is transformed into the energy of internal (microscopic molecular) motion in friction and impact.

Let us now show the form to which the total energy reduces when the Lagrangian can be represented in the form $L = T - U$, where the kinetic energy T is a homogeneous quadratic function of the generalized velocities, $T = (1/2) T_{\alpha\beta} \dot{q}_\alpha \dot{q}_\beta$. Since we have assumed that the potential energy depends only on the coordinates, for the derivatives with respect to the generalized velocities we have

$$\frac{\partial L}{\partial \dot{q}_\alpha} = \frac{\partial T}{\partial \dot{q}_\alpha}$$

so that the total energy is

$$E = \dot{q}_\alpha \frac{\partial T}{\partial \dot{q}_\alpha} - L \tag{4.3}$$

But according to Euler's theorem on homogeneous functions the sum of the partial derivatives multiplied by the corresponding variables is equal to the function itself multiplied by the degree of homogeneity (this is easily verified from the function of two variables $ax^2 + 2bxy + cy^2$). Hence

$$E = 2T - (T - U) = T + U \tag{4.4}$$

that is, the total energy is equal to the sum of the potential and kinetic energies, in agreement with the conventional definition. This explains the names given to the functions T and U.

Note that the definition of energy (4.1) is more general and can be used in the case when the Lagrangian cannot be represented as $L = T - U$ (see Part II).

Application of the Energy Integral to Systems with One Degree of Freedom. The energy integral allows us, straightaway, to reduce problems of the motion of systems with one degree of freedom to those of quadrature. Thus, in the pendulum problem considered in the previous section we can, with the aid of (4.4), write the energy integral directly:

$$E = \frac{m}{2} l^2 \dot{\varphi}^2 + mgl (1 - \cos \varphi) \tag{4.5}$$

The value of E is determined from the initial conditions. For example, let the pendulum initially be deflected at an angle φ_0 and released without any initial speed: so $\dot{\varphi}_0 = 0$. Hence

$$E = mgl (1 - \cos \varphi_0)$$

Substituting this into (4.5), we have

$$mgl \, (\cos \varphi - \cos \varphi_0) = \frac{m}{2} \, l^2 \dot{\varphi}^2 \tag{4.6}$$

From this, the relationship between the angle of deflection and time is determined by the quadrature

$$t = -\left(\frac{l}{2g}\right)^{1/2} \int\limits_{\varphi_0}^{\varphi} \frac{d\varphi}{(\cos \varphi - \cos \varphi_0)^{1/2}} \tag{4.7}$$

The minus sign has been taken because at the beginning of the motion the angle φ decreases. The integral involved in (4.7) cannot be found explicitly.

It is significant that the oscillation law of a pendulum depends only on the ratio l/g. The mass, as pointed out in the preceding section, is eliminated. Thus, a pendulum can be used to measure the acceleration of free fall, g, to a high degree of accuracy.

A system in which mechanical energy is conserved is sometimes called a *conservative system*. Thus, the energy integral makes it possible to reduce the problem of the motion of a conservative system with one degree of freedom to quadratures. The fact that the quadrature need not necessarily be expressed in terms of elementary functions, as is the case in (4.7), is of no consequence.

In a conservative system with several degrees of freedom the energy integral allows us to reduce the order of the set of differential equations by one unit and thereby simplify the integration.

Generalized Momentum. We shall now consider other integrals of motion which can be found directly with the aid of the Lagrangian. To do this we shall take advantage of the following, quite obvious, consequence of the Lagrange equations. If some coordinate does not appear explicitly in the Lagrangian, $\partial L / \partial q_\alpha = 0$, then in accordance with the Lagrange equations

$$\frac{d}{dt} \frac{\partial L}{\partial \dot{q}_\alpha} = 0 \tag{4.8}$$

But then

$$p_\alpha \equiv \frac{\partial L}{\partial \dot{q}_\alpha} \tag{4.9}$$

is constant, which implies that it is an integral of the motion. The quantity p_α is called the *generalized momentum* corresponding to the coordinate with index α. This definition includes the momentum in the usual sense

$$p_x = mv_x = \frac{\partial L}{\partial v_x} = \frac{\partial L}{\partial \dot{x}} \tag{4.10}$$

Summarizing, if a certain generalized coordinate does not appear explicitly in the Lagrangian, the generalized momentum corresponding to it is an integral of the motion, that is, it remains constant during the motion. The coordinate itself is said to be *cyclic*.

In the preceding section we saw that the coordinates X, Y, Z of the centre of mass of a system of two particles not subject to external forces do not appear in the Lagrangian. From this it is evident that

$$(m_1 + m_2) \dot{X} = P_X$$

$$(m_1 + m_2) \dot{Y} = P_Y$$

$$(m_1 + m_2) \dot{Z} = P_Z \tag{4.11}$$

are integrals of the motion.

Momentum of a System of Particles. The same is readily shown also for a system of N particles. Indeed, for N particles we can introduce the concept of centre of mass by means of the equation

$$\mathbf{R} = \sum_i m_i \mathbf{r}_i / \sum_i m_i \tag{4.12}$$

and of the velocity of the centre of mass as

$$\dot{\mathbf{R}} = \sum_i m_i \dot{\mathbf{r}}_i / \sum_i m_i \tag{4.13}$$

The velocity of the ith point relative to the centre of mass is (by the theorem of the composition of velocities)

$$\dot{\mathbf{r}}_i' = \dot{\mathbf{r}}_i - \dot{\mathbf{R}} \tag{4.14}$$

The kinetic energy of the system of particles is

$$T = \frac{1}{2} \sum_{i=1}^{N} m_i \dot{\mathbf{r}}_i^2 = \frac{1}{2} \sum_{i=1}^{N} m_i (\dot{\mathbf{r}}_i' + \dot{\mathbf{R}})^2$$

$$= \frac{1}{2} \sum_{i=1}^{N} m_i \dot{\mathbf{r}}_i'^2 + \dot{\mathbf{R}} \sum_{i=1}^{N} m_i \dot{\mathbf{r}}_i' + \frac{1}{2} \sum_{i=1}^{N} m_i \dot{\mathbf{R}}^2 \tag{4.15}$$

But from (4.13) and (4.14) it is immediately apparent that, from the definition of $\mathbf{r}_i'$ and $\mathbf{R}$, $\sum_{i=1}^{N} m_i \dot{\mathbf{r}}_i' = 0$. Therefore the kinetic energy of a system of particles separates into a sum of two terms: the kinetic energy of motion of the total mass with the velocity of the centre of mass relative to the adopted reference frame, and the kinetic

energy of motion of the masses relative to the centre of mass:

$$T = T_{c.m} + T_{rel} = \frac{1}{2} \Big(\sum_{i=1}^{N} m_i \Big) \dot{\mathbf{R}}^2 + \frac{1}{2} \sum_{i=1}^{N} m_i \dot{\mathbf{r}}_i'^2 \qquad (4.16)$$

The vectors $\dot{\mathbf{r}}_i'$ are not independent: as mentioned, they are governed by the vector equation $\sum m_i \mathbf{r}_i' = 0$. Consequently they can be expressed in terms of $N-1$ independent quantities by defining the relative positions of the ith point and some fixed point, for instance the first. The kinetic energy of relative motion is expressed in terms of the relative velocities of the particles (see Exercise 4, Section 3). In a closed system, by virtue of its fundamental property, there are no external forces acting on the particles, while the forces of interaction within the system depend only on the relative configuration of the particles, that is, on $\mathbf{r}_i - \mathbf{r}_k$.

Thus, only $\dot{\mathbf{R}}$ appears in the Lagrangian, and $\mathbf{R}$ does not. Its components are cyclic coordinates, therefore the total momentum is conserved

$$\mathbf{P} = \frac{\partial L}{\partial \dot{\mathbf{R}}} = \Big(\sum_{i}^{N} m_i \Big) \dot{\mathbf{R}} = \sum_{i=1}^{N} m_i \dot{\mathbf{r}}_i \qquad (4.17)$$

Equation (4.17) shows that the total momentum of a system of mass points not subject to the action of external forces is an integral of the motion. It is important that this is an additive integral of the motion compounded from the momenta of individual particles.

Note that the momentum integral exists for any system subject only to internal forces, including friction, which cause mechanical energy to transform into the energy of internal molecular motion. This does not affect the conservation of momentum.

If (4.17) is again integrated with respect to time the result is a centre-of-mass integral analogous to (3.18). This is the so-called *second integral* (since it involves two constants). It contains only the current coordinates, but not the velocities; (4.7) is also a second integral.

Properties of the Vector Product. Further on we shall investigate the *moment of momentum*, or *angular momentum*, of a mass point and a system of mass points. For a separate mass point it is defined as

$$\mathbf{M} = \mathbf{r} \times \mathbf{p} \qquad (4.18)$$

Here the boldface multiplication sign denotes the vector product. As is known, (4.18) takes the place of the following three equations

for the components:

$$M_x = yp_z - zp_y$$
$$M_y = zp_x - xp_z$$
$$M_z = xp_y - yp_x \tag{4.19}$$

Recalling the geometric definition of a vector product, we construct a parallelogram on vectors **r** and **p**. Then **r** ✕ **p** denotes a vector equal in magnitude to the area of the parallelogram and directed perpendicular to its plane. To state the direction of **M** uniquely we must agree on the direction in which we trace the parallelogram. We shall always start from the first vector, in this case **r**. Then the positive side of the parallelogram is that in which the direction is counterclockwise. Vector **M** is normal to the positive side of the plane. To state this in another way, if we rotate a corkscrew from **r** to **p**, the displacement of the corkscrew itself will be in the direction of **M**. The direction changes if we interchange the positions of **r** and **p** in the product. Therefore, unlike a conventional product, the sign of the vector product changes when the factors are interchanged. This can also be seen from the definition of the Cartesian components of angular momentum.

In order to understand why a vector product defines precisely a vector quantity we should clarify the definition of a vector in general. A vector is an aggregate of three quantities which transform in the rotations of a coordinate system as the components of the radius vector. For example, velocity is a vector, because by definition it is dr/dt, and differentials of coordinates transform like the coordinates themselves. Consequently, momentum **p** is also a vector.

If we carry out the transformation of the components of the radius vector (x, y, z) and of the momentum (p_x, p_y, p_z) according to the formulas of analytic geometry and substitute the transformed quantities into (4.19), we find that the components M_x, M_y, M_z have themselves transformed according to the same formulas as x, y, z. For this we should make use of the known relationships for the cosines of the angles between the old and new coordinate axes. But if some three equations, in this case (4.19), retain their form in rotations of the coordinate system, they can be combined in one vector equation (4.18) (Exercise 3).

The area of the parallelogram is $rp \sin \alpha$, where α is the angle between **r** and **p**. The product $r \sin \alpha$ is the length of a perpendicular drawn from the origin of the coordinate system to the tangent to the trajectory whose direction is the same as **p**. This length is sometimes called the "arm" of the moment.

The vector product possesses a distributive property, that is

$$\mathbf{a} \times (\mathbf{b} + \mathbf{c}) = (\mathbf{a} \times \mathbf{b}) + (\mathbf{a} \times \mathbf{c})$$

Hence, a binomial vector product is calculated in the usual way, but the order of the factors is taken into account:

$$(\mathbf{a} + \mathbf{b}) \times (\mathbf{c} + \mathbf{d}) = (\mathbf{a} \times \mathbf{c}) + (\mathbf{b} \times \mathbf{c}) + (\mathbf{a} \times \mathbf{d}) + (\mathbf{b} \times \mathbf{d})$$

Angular Momentum of a System of Mass Points. The angular momentum of a system of mass points is defined as the sum of the angular momenta of all the points taken separately. In doing so we must, of course, take the radius vectors relative to a coordinate origin common to all the particles. Thus

$$\mathbf{M} = \sum_{i=1}^{N} \mathbf{r}_i \times \mathbf{p}_i \tag{4.20}$$

We shall show that the angular momentum of a system can be separated into the angular momentum of relative motion of the mass points and the angular momentum of the system as a whole, similar to the way it was done for the kinetic energy. To do this we must represent the radius vector of each mass point as the sum of the radius vector of its position relative to the centre of mass and the radius vector of the centre of mass itself; we must expand the expression for the velocities of the mass points in the same way. Then, the angular momentum can be written in the form

$$\mathbf{M} = \sum_{i=1}^{N} (\mathbf{R} + \mathbf{r}'_i) \times (m_i \dot{\mathbf{R}} + \mathbf{p}'_i)$$
$$= \sum_{i=1}^{N} [(m_i \mathbf{R} \times \dot{\mathbf{R}}) + (m_i \mathbf{r}'_i \times \dot{\mathbf{R}}) + (\mathbf{R} \times \mathbf{p}'_i) + (\mathbf{r}'_i \times \mathbf{p}'_i)]$$

In the second and third terms, we can make use of the distributive property of the vector product and introduce the summation sign inside the parentheses. However, both these sums are equal to zero, by definition of the centre of mass. This was used in (4.15) for velocities. Thus, the angular momentum of a system of mass points is indeed equal to the sum of the angular momenta of the centre of mass, $\mathbf{M}_0$, and the relative motion of the mass points, $\mathbf{M}'$:

$$\mathbf{M} = \mathbf{R} \times \mathbf{P} + \sum_i \mathbf{r}'_i \times \mathbf{p}'_i \equiv \mathbf{M}_0 + \mathbf{M}' \tag{4.21}$$

Let us perform these transformations for the special case of a system of two mass points. We substitute $\mathbf{r}_1$ and $\mathbf{r}_2$ expressed from (3.14) in terms of $\mathbf{r}$ and $\mathbf{R}$ into (4.20). This gives

$$\mathbf{M} = \mathbf{r}_1 \times \mathbf{p}_1 + \mathbf{r}_2 \times \mathbf{p}_2$$
$$= \mathbf{R} \times (\mathbf{p}_1 + \mathbf{p}_2) + \frac{1}{m_1 + m_2} (m_2 \mathbf{r} \times \mathbf{p}_1 - m_1 \mathbf{r} \times \mathbf{p}_2)$$

Further, we replace $\mathbf{p}_1$ by $m_1\dot{\mathbf{r}}_1$, $\mathbf{p}_2$ by $m_2\dot{\mathbf{r}}_2$, and $\mathbf{p}_1 + \mathbf{p}_2$ by $\mathbf{P}$, after which the angular momentum reduces to the required form:

$$\mathbf{M} = \mathbf{R} \times \mathbf{P} + \frac{m_1 m_2}{m_1 + m_2}\, \mathbf{r} \times \dot{\mathbf{r}} \tag{4.22}$$

Here, $[m_1 m_2/(m_1 + m_2)]\,\dot{\mathbf{r}} = m\dot{\mathbf{r}} = \mathbf{p}$ is the momentum of relative motion of the mass points. It involves the reduced mass m.

We shall now show that the angular momentum of relative motion does not depend on the choice of the origin. Indeed, if we displace the origin, then all the quantities $\mathbf{r}_i'$ change by the same amount: $\mathbf{r}_i' = \mathbf{r}_i'' + \mathbf{a}$.

Accordingly, the angular momentum of relative motion will be

$$\mathbf{M}' = \sum_{i=1}^{N} \mathbf{r}_i' \times \mathbf{p}_i' = \sum_{i=1}^{N} \mathbf{r}_i'' \times \mathbf{p}_i' + \sum_{i=1}^{N} \mathbf{a} \times \mathbf{p}_i'$$

$$= \sum_{i=1}^{N} \mathbf{r}_i'' \times \mathbf{p}_i' + \mathbf{a} \times \sum_{i=1}^{N} \mathbf{p}_i' = \mathbf{M}''$$

because

$$\sum_{i=1}^{N} \mathbf{p}_i' = \sum_{i=1}^{N} m_i \dot{\mathbf{r}}_i' = 0$$

Thus, in calculating the angular momentum of relative motion the origin of the coordinates can be placed at any point.

Conservation of Angular Momentum. We shall now show that the angular momentum of an isolated system is an integral of the motion. First take the total angular momentum of the system. Its time derivative is

$$\frac{d\mathbf{M}_0}{dt} = \dot{\mathbf{R}} \times \mathbf{P} + \mathbf{R} \times \dot{\mathbf{P}} = 0$$

because $\dot{\mathbf{P}} = 0$ for systems not subject to external forces, and $\dot{\mathbf{R}} \times \mathbf{P} = 0$ because $\dot{\mathbf{R}}$ is directed along $\mathbf{P}$.

Let us prove that the angular momentum of the relative motion of the mass points in the system is also conserved. Its total derivative with respect to time is

$$\frac{d\mathbf{M}'}{dt} = \sum_{i=1}^{N} \dot{\mathbf{r}}_i' \times \mathbf{p}_i' + \sum_{i=1}^{N} \mathbf{r}_i' \times \dot{\mathbf{p}}_i' \tag{4.23}$$

The first term in the right-hand side of (4.23) vanishes because $\dot{\mathbf{r}}_i'$ is directed along $\mathbf{p}_i'$. Let us consider the second term. Recalling that the potential energy U depends on the absolute values of the differ-

ences of all the coordinates, $U = U (\ldots | \mathbf{r}_i - \mathbf{r}_k | \ldots)$, from the equations of motion we can write, with the help of formulas (3.1) and (3.2)

$$\frac{d\mathbf{p}'_i}{dt} = -\sum_{k \neq i} \frac{\partial U}{\partial | \mathbf{r}_{ik} |} \frac{\mathbf{r}_{ik}}{| \mathbf{r}_{ik} |}$$

Here in the right-hand side we have the sum of all the forces with which all the other points act on the ith point. We derive the vector product of this equation multiplied by $\mathbf{r}'_i$ and sum over all the mass points. On the left we have the time derivative of the angular momentum of the relative motion of the system, and on the right, the double sum over all pairs. The partial derivatives $\partial U/\partial | \mathbf{r}_{ik} |$ are involved in this sum twice: of the ith and kth mass points. They are multiplied by the vectors $\mathbf{r}_i \times (\mathbf{r}_i - \mathbf{r}_k)$ and $\mathbf{r}_k \times (\mathbf{r}_k - \mathbf{r}_i)$, which in the sum vanish because $\mathbf{r}_i \times \mathbf{r}_i = 0$, $\mathbf{r}_k \times \mathbf{r}_k = 0$, $\mathbf{r}_i \times \mathbf{r}_k = -\mathbf{r}_k \times \mathbf{r}_i$. Thus $d\mathbf{M}'/dt = 0$, so that the angular momentum of the relative motion of the mass points in a closed system is conserved.

In the next section we shall show that angular momentum (or its components) may be conserved in an external field as well, provided the field possesses the required symmetry.

Additive Integrals of Motion for a Closed System. We have shown that a closed mechanical system has the following first integrals of motion: energy, three components of the momentum vector, and three components of the angular momentum vector. Linear and angular momenta are always additive, and energy is additive only for the noninteracting parts of the system.

The mechanical energy referred to macroscopic degrees of freedom of a body as a whole is in very many cases not conserved. In the presence of friction forces it is transferred in the form of heat to the microscopic (molecular) degrees of freedom. Linear and angular momenta are always conserved in a closed mechanical system. The former is associated with the motion of the system's centre of mass, the latter with its rotation about the centre of mass. Both these integrals of the motion belong to the macroscopic (mechanical) degrees of freedom.

It is much more difficult to obtain all other mechanical integrals of motion (with the exception of the centre-of-mass integral), and no general rule for determining them can be formulated.

The seven additive integrals of motion—energy, linear and angular momenta (seven because the latter two are vectors)—are special cases in the sense that they owe their existence to symmetry with respect to translations and rotations. Indeed, the symmetry of the Lagrangian with respect to displacements in time leads to the energy

conservation law. Symmetry with respect to spatial displacements imposes the restriction on the potential energy that it depend only on the differences between the particles' coordinates. Thanks to this the motion of the centre of mass is separated, and the total momentum of the system is conserved. We deduced the conservation of the angular momentum of relative motion from the fact that the potential energy involves the absolute values of the differences $|\mathbf{r}_i - \mathbf{r}_k|$, which is in agreement with the property of space isotropy. From this property can be deduced the conservation of angular momentum in the case of less restrictive assumptions regarding U (Sec. 5).

EXERCISES

1. Describe the motion of a mass point moving along a cycloid in a gravitational field.

Solution. The equation of the cycloid in parametric form is

$$z = -R \cos s, \qquad x = Rs + R \sin s$$

The kinetic energy of the point is

$$T = \frac{m}{2}(\dot{x}^2 + \dot{z}^2) = 2mR^2 \cos^2 \frac{s}{2} \dot{s}^2$$

The potential energy is

$$U = -mgR \cos s$$

The total-energy integral is

$$E = 2mR^2 \cos^2 \frac{s}{2} \dot{s}^2 - mgR \cos s = \text{constant}$$

The value of E can be determined on the condition that the velocity $\dot{s}$ is equal to zero when the deflection is maximum, $s = s_0$; the mass point moves along the cycloid from that position. Hence

$$E = -mgR \cos s_0$$

After separating the variables and integrating, we obtain

$$t = \sqrt{2}\,R \int \frac{\cos(s/2)\,ds}{(gR\cos s + E/m)^{1/2}} = \left(\frac{2R}{g}\right)^{1/2} \int \frac{\cos(s/2)\,ds}{(\cos s - \cos s_0)^{1/2}}$$

Denoting $\sin(s/2) = u$, we integrate and put the limits to get

$$t = \left(\frac{R}{g}\right)^{1/2} \int \frac{2\,du}{(u_0^2 - u^2)^{1/2}} = 2\left(\frac{R}{g}\right)^{1/2} \arcsin \frac{u}{u_0}$$

In order to find the total period of the motion, we must take the integral between the limits $-u_0$ and $+u_0$ and double the result. This corresponds to the oscillation of the mass point from $s = -s_0$ to $s = s_0$, and back to $s = -s_0$.

Thus the total period of oscillation is equal to $2\pi(R/g)^{1/2}$.   ng as the mass point moves on the cycloid, the period of its oscillation does not depend on the oscillation amplitude s_0 (Huygens' cycloidal pendulum). The period of oscillation of an ordinary pendulum, which describes an arc of a circle, is known, in the general case, to depend on the amplitude (cf. (4.7))•

2. Prove that a point moving along a curved line in a vertical plane descends from a given upper position to a given lower position in the shortest time if the curve is a cycloid.

Solution. Using the results obtained in Exercise 2, Section 3, we write the energy integral for the point descending along the curve:

$$E = \frac{m}{2}!(x'^2 + z'^2)\left(\frac{ds}{dt}\right)^2 + mgz\,(s)$$

Hence the time of descent is given by the quadrature

$$t = g^{-1/2} \int \frac{ds\,(x'^2 + z'^2)^{1/2}}{[z(s_0) - z(s)]^{1/2}}$$

Passing in the integral from the independent variable s to the variable z we obtain

$$t = g^{-1/2} \int \frac{dz\,[1 + (dx/dz)^2]^{1/2}}{(z_0 - z)^{1/2}}$$

The dependence of x on z must be so defined as to assure that t has an extremum. We considered a similar problem in connection with Hamilton s principle, where we had to find the path corresponding to the minimum value of the integral S. For this the integrand must satisfy the Lagrange equation. Such an equation can, obviously, be written for solving the present problem, taking z as the independent variable, x as the dependent variable, and the integrand taken from the expression for t. Since the dependent variable is not explicitly involved in the integrand, it is cyclic, and the corresponding "momentum", that is, the derivative of the integrand with respect to dz, is constant. Denoting it $-(2R)^{-1/2}$, we have

$$\frac{dx}{dz}\left[1 + \left(\frac{dx}{dz}\right)^2\right]^{-1/2} = -\left(\frac{z_0 - z}{2R}\right)^{1/2}$$

Introducing new variables

$$z = z_0 - R\,(1 + \cos s), \qquad dz = R\sin s\,ds$$

yields

$$\frac{dx}{dz} = \cot \frac{s}{2} \qquad\qquad\qquad .$$

Passing from z to s, we easily find

$$dx = R\,(1 + \cos s)\,ds, \qquad x = R\,(s + \sin s)$$

All that is needed to obtain the cycloid equation written in the previous exercise is to put $z_0 = R$.

From Exercise 1 the total time in which the particle descends from the upper to the lower position is equal to $(\pi/2)(R/g)^{1/2}$. In free fall from the height $2R$, equal to the ascent of the cycloid, the time is $2(R/g)^{1/2}$.

3. Show that in rotations of the coordinates the vector product defined by formulas (4.19) transforms as a vector.

Solution. The transformation formulas for the components of a vector in a rotation of a coordinate system are known from analytic geometry. We write these formulas in compact form. Number the coordinates $x = x_1$, $y = x_2$, $z = x_3$ and denote the cosine of the angle between the new axis α and the old axis β as (α, β). Then the required form of the formulas is (taking the sum with respect to the index that appears twice)

$$x'_\alpha = (\alpha,\ \beta)\, x_\beta$$

In view of symmetry between the old and new coordinates the inverse transformation is written as follows:

$$x_\beta = x'_\alpha\,(\alpha,\ \beta) \quad \text{or} \quad x_\alpha = x'_\beta\,(\beta,\ \alpha)$$

The determinant of the transformation is unity. This is proved in the following way. As is known from analysis, in a transformation of coordinates the volume element is multiplied by the Jacobian determinant

$$\frac{\partial\,(x'_1,\ x'_2,\ x'_3)}{\partial\,(x_1,\ x_2,\ x_3)}$$

In this case the Jacobian coincides with the transformation determinant, because it is linear. On the other hand, a rotation cannot affect the volume, so that the assertion is proved.

Now we construct the determinant of the inverse transformation in general form. It is known that the coefficient at the intersection of row α and column β in an inverse transformation is the minor of the corresponding element in the direct transformation divided by the determinant, in the present case by unity. At the same time, the inverse transformation is carried out with the help of the same coefficients but with the indices interchanged. We now write the determinant in explicit form:

$$\begin{vmatrix} (1,1) & (1,2) & (1,3) \\ (2,1) & (2,2) & (2,3) \\ (3,1) & (3,2) & (3,3) \end{vmatrix}$$

Equating the elements of the first row with the interchanged indices to the corresponding minors, we obtain

$$(1,1) = (2,2)\,(3,3) - (2,3)\,(3,2)$$
$$(2,1) = (2,3)\,(3,1) - (2,1)\,(3,3)$$
$$(3,1) = (2,1)\,(3,2) - (3,1)\,(2,2)$$

Now take three unit vectors directed along the axes of a right-handed coordinate system (rotation of the corkscrew handle from x to y displaces it along the z axis). From the definition of the vector product the following

equation must hold:

$$\mathbf{i} = \mathbf{j} \times \mathbf{k}$$

where the unit vector **i** is directed along x, **j** is along y, and **k** is along z.

For the equation actually to occur it should be satisfied not only in the initial coordinate system. Projecting the unit vectors on new axes, we see that the obtained relationships for the cosines of the angles between the old and new axes assure the validity of the relationship between vectors **i**, **j**, **k** in all systems.

Since any two vectors can be resolved along unit vectors and a vector product is distributive, it always yields a vector. Note that this is in a sense a "fortuitous" property of vectors in three-dimensional space, insofar as there happen to be three quantities of the type found in the right-hand sides of (4.19).

5

MOTION IN A CENTRAL FIELD

The Angular Momentum Integral. We shall consider the motion of two bodies (particles) in a frame of reference fixed with respect to their centre of mass. In such a reference frame angular momentum is associated only with the relative motion of the bodies. If there are no external forces the angular momentum is conserved. We denote the separation of the particles $\mathbf{r} = \mathbf{r}_1 - \mathbf{r}_2$; the corresponding linear momentum is

$$\mathbf{p} = \frac{m_1 m_2}{m_1 + m_2}\, \dot{\mathbf{r}} \equiv m\mathbf{v} \tag{5.1}$$

This relationship is obtained from (3.17) by differentiation of the particles' Lagrangian with respect to $\dot{\mathbf{r}} \equiv \mathbf{v}$, whence in accordance with (4.22) we find

$$\mathbf{M} = \mathbf{r} \times \mathbf{p} = m\mathbf{r} \times \dot{\mathbf{r}} = \text{constant} \tag{5.2}$$

If the vector is constant, all three components are constant. Not only its absolute value but its spatial direction as well does not change. But from (5.2) the vector **r** of the relative position of the particles and the vector $\dot{\mathbf{r}}$ of their relative velocity are perpendicular to the constant direction of **M**. And since all perpendiculars to the same point of a line lie in one plane, the relative motion of the particles of the system takes place in that plane. It can be seen from

(5.2) that vector **r** lies in the same plane as the path, which is a plane curve.

When transforming to a spherical coordinate system, it is advisable to direct the polar axis along z. Then the motion takes place in the x,y-plane, or $\vartheta = \pi/2$, $\sin \vartheta = 1$, $\dot\vartheta = 0$.

The potential energy of a system of two mass points depends only on the distance r between them, because it is the only scalar quantity that can be derived from vector **r**. From (3.11) the Lagrangian for plane motion at $\dot\vartheta = 0$, $\sin \vartheta = 1$ is

$$L = \frac{m}{2} (\dot{r}^2 + r^2\dot\varphi^2) - U(r) \tag{5.3}$$

where m is the reduced mass.

If one mass is much greater than the other, it lies very close to the centre of mass of the system and its motion can be neglected. In that case the light particle (or system of light particles) moves around the heavy one like a planet in the solar system. The angular momentum conservation law holds, but only with respect to the central body and not to an arbitrary point in space. The conclusion that the paths of the moving particles are plane also holds, but only provided their interactions can be neglected or that they were moving in the same plane from the outset.

Angular Momentum as Generalized Momentum. We shall now show that the angular momentum component M_z, which in the adopted coordinate system is equal simply to the absolute value M, is nothing other than $p_\varphi = \partial L/\partial\dot\varphi$, that is, the generalized linear momentum corresponding to the coordinate φ, which is the angle of rotation about the z axis. Indeed, from (4.19), (3.8b), and (3.8c), the angular momentum $M = M_z$ is

$$M = M_z = xp_y - yp_x = mr\cos\varphi\,(\dot{r}\sin\varphi + r\cos\varphi\,\dot\varphi)$$

$$- mr\sin\varphi\,(\dot{r}\cos\varphi - r\sin\varphi\,\dot\varphi)$$

$$= mr^2(\cos^2\varphi + \sin^2\varphi)\,\dot\varphi = mr^2\dot\varphi$$

On the other hand, differentiating L with respect to $\dot\varphi$ we see that

$$p_\varphi = \frac{\partial L}{\partial\dot\varphi} = mr^2\dot\varphi \tag{5.4}$$

The quantity $r^2\dot\varphi/2$ is known as the *areal velocity*, that is, the area swept out by the radius vector of a mass point in unit time. Indeed, $r\dot\varphi$ is the base of a triangle whose vertex is at the origin, and r is its altitude. The difference between the areas of this triangle and

the triangle produced by the displacement of the particle, which need not be perpendicular to the radius, is an infinitesimal of the second order. Thus, in geometrical interpretation the angular momentum conservation law expresses the constancy of the areal velocity in the orbital motion of a material point in a field of central forces (*Kepler's Second Law*).

The relationship $M_z = p_\varphi$ offers a new explanation of the conservation of angular momentum of a closed system. Indeed, the Lagrangian of such a system cannot change in a rotation of the frame of reference through an arbitrary angle about an arbitrary axis in space. But then the angle of rotation is a cyclic coordinate, and the corresponding generalized momentum is conserved. And since the angle of rotation is arbitrary, all three angular momentum components of the closed system must be conserved. It can be seen from this reasoning that the law of conservation of angular momentum holds not only when the forces between points act along the lines joining them (as was assumed in Section 4) but in the most general case as well. Direct proof of this on the basis of the equations of motion is rather cumbersome.

Elimination of the Azimuthal Velocity Component. The angular momentum integral permits us to reduce the problem of two-particle motion, or the problem of motion of a single particle in a central field, to quadrature. To do this we must express $\dot\varphi$ in terms of angular momentum and thus get rid of the superfluous variable, since angle φ itself does not appear in the Lagrangian. In this fashion every cyclic variable can be eliminated.

In accordance with (4.4), we first of all have the energy integral

$$E = \frac{m}{2}\,(\dot r^2 + r^2\dot\varphi^2) + U\,(r) \tag{5.5}$$

Eliminating $\dot\varphi$ with the aid of (5.4), we obtain

$$E = \frac{m\dot r^2}{2} + \frac{M^2}{2mr^2} + U\,(r) = \text{constant} \tag{5.6}$$

This first-order differential equation (in r) is later reduced to quadrature. Before writing down the quadrature, let us examine it graphically.

The Dependence of the Form of Path on the Sign of the Energy. For such an examination, we must make certain assumptions about the variation of potential energy.

From (2.1), force is connected with potential energy by the relation

$$\mathbf{F} = -\frac{\partial U}{\partial \mathbf{r}}\,, \qquad U = \int\limits_r F\,dr$$

The upper limit in the integral can be chosen arbitrarily. If F tends to zero at infinite r faster than $1/r$, then the integral $\int_r^\infty F\,dr$ is convergent. Then we can put $U(r) = \int_r^\infty F\,dr$, or $U(\infty) = 0$. In other words, the potential energy is chosen to be zero at infinity.

Furthermore, we assume that $U(r)$ does not increase faster than $1/r$ at $r \to 0$, as, for example, in Newtonian attraction, where $U = -\int_r^\infty (a/r^2)\,dr = -a/r$.

Let us now write (5.6) as

$$\frac{m\dot{r}^2}{2} = E - \frac{M^2}{2mr^2} - U(r) \qquad (5.7)$$

The left-hand side of this equation is always positive. At $r \to \infty$ the last two terms tend to zero. Thus for the **particles to be able** to recede from each other to an infinite distance, the total energy must be positive when the potential energy satisfies the condition $U(\infty) = 0$. Hence, if two particles come together from infinity, their energy must, according to the conservation law, be positive. If, as the two particles draw closer, the energy is not transferred to a third particle, on meeting they will of necessity separate again to an infinite distance.

Given a definite form of U, we can plot the curve of the function

$$U_M(r) \equiv \frac{M^2}{2mr^2} + U(r) \qquad (5.8)$$

The index "M" denotes that the potential energy includes the "centrifugal" energy $M^2/(2mr^2)$. The derivative of this quantity with respect to r, taken with the opposite sign, is equal to $M^2/(mr^3)$. If we put $M = mr^2\dot{\varphi}$, the result will be the usual expression for the centrifugal force. However, in future we shall call a mechanical quantity of different origin the "centrifugal force" (Sec. 8).

Let $U(r) < 0$ and increase monotonically as r changes from zero to infinity. It follows that the force has a negative sign $\left(\text{since } \mathbf{F} = -\frac{\partial U}{\partial \mathbf{r}}\right)$, that is, it is an attractive force. Let us assume, in addition, that at infinity $|U(r)| > \frac{M^2}{(2mr^2)}$. This is true, for example, of Newtonian gravitational forces or for electrostatic Coulomb forces between charged bodies.

Let **us** summarize the assumptions we have made regarding $U_M(r)$:

(i) at $r \to 0$ the "centrifugal" term is predominant, hence $U_M(r)$ is infinite and positive;

(ii) at $r \to \infty$, where $|U(r)| > M^2/(2mr^2)$, $U_M(r)$ tends to zero from the side of negative values.

Consequently, the curve $U_M(r)$ has the form shown in Figure 5. On the side of small values of r it decreases away from zero as $1/r^2$;
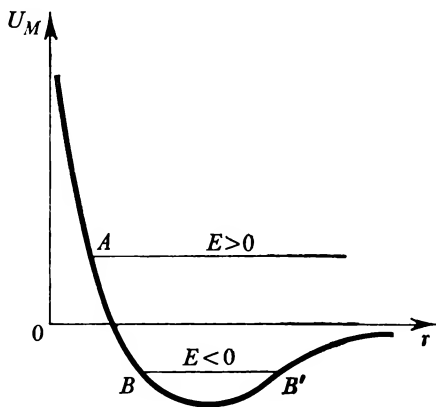

Figure 5

as it approaches large values of r it increases as $-a/r$, approaching the r axis from below. The curve must have a minimum in the domain of median values of r.

In the interaction of a charged body with a neutral one (for instance, with an atom which has not lost a single electron) $U(r)$ decreases faster than the "centrifugal" energy. Therefore at large distances $U_M(r)$ approaches the r axis from the positive side. Then the curve $U_M(r)$ first passes through a minimum, then increases, and after passing a maximum decreases again, tending to zero as $r \to \infty$. This holds if there is a domain where $U(r)$ predominates over the "centrifugal" energy. Otherwise the decrease is monotonic.

The total energy of a system of converging particles can also be plotted in Figure 5. Since E is conserved in motion, the curve has the form of a horizontal straight line lying above or below the z axis, depending on the sign of E. For positive values of energy, the line $E =$ constant lies above the curve $U_M(r)$ everywhere to the right of point A. In this case the difference $E - U_M(r)$ is positive. The particles can approach each other from infinity and recede from each other to infinity. Such motion is termed *infinite*. As we shall see

later in this section, in the case of Newtonian attraction we obtain hyperbolic orbits.

At $E < 0$, but higher than the minimum of the curve $U_M(r)$, the difference $E - U_M(r) = m\dot{r}^2/2$ remains positive only between points B and B'. Consequently, between the corresponding two values of the radius lies a physically possible domain of motion with the given negative total energy. The motion is in this case termed *finite*. In the case of Newtonian attraction elliptical orbits correspond to it. In the motion of a planet about the sun point B corresponds to the perihelion, point B' to the aphelion.

At $E = 0$ the motion is infinite. As r increases the velocity tends to zero, remaining positive. In the case of Newtonian attraction parabolic orbits correspond to this value of E.

Falling Onto the Centre. From the preceding reasoning it is apparent that r cannot decrease to zero owing to "centrifugal" energy. Only if the particles are "targeted" on each other does the arm vanish, so that $M = 0$, and the curve $U_M(r)$ is replaced by the curve $U(r)$. Then nothing prevents the particles from colliding.

Let us now investigate an imaginary rather than a real case, when $-U(r)$ tends to infinity as $r \to 0$ faster than $1/r^2$. In this case $U_M(r)$ is negative for all r's close to zero. From (5.7), $\dot{r}^2$ is positive at infinitesimal values of r and tends to infinity as $r \to 0$. But the particle cannot collide head-on because such a collision would violate the law of conservation of angular momentum. The angular momentum is equal to $m\rho v_\varphi$, where ρ is the arm. For M to retain its given finite value when the arm decreases to infinitesimal values the velocity component v_φ perpendicular to the radius must tend to infinity as $1/\rho$. Then the product $m\rho v_\varphi$, which defines the angular momentum, remains finite.

Thus, if $U_M = -\infty$ at $r = 0$, the radial component of the velocity tends to zero, and the azimuthal component tends to infinity. The path of the particle has the form of a helix winding around the attracting centre, but never reaching it. The coils of the helix decrease, but the rotation speed increases. The force of "centrifugal" repulsion cannot prevent the particles from drawing gradually closer, which takes place the slower the smaller r is.

In the motion of three bodies gravitating towards one another according to Newton's law, two of them may collide even if at the initial time their motion was not purely radial. Indeed, only the total angular momentum of relative motion is conserved, and this does not preclude the collision of the two bodies.

Finally, for repulsive forces tending to infinity as $r \to 0$, falling onto the centre is impossible. Obviously in this case the motion is only infinite.

Reducing to Quadrature. Let us now find the equation of the path in general form. To do this we must in (5.7) change from differentiation with respect to time to differentiation with respect to φ. Using (5.4) we have

$$dt = \frac{mr^2}{M} d\varphi \tag{5.9}$$

Separating the variables and passing to φ in (5.7) yields

$$\varphi = \int_{r_0}^{r} \frac{M}{mr^2} \left[\frac{2}{m} \left(E - \frac{M^2}{2mr^2} - U(r) \right) \right]^{-1/2} dr \tag{5.10}$$

As can be seen from the equation, angle φ is put equal to zero at $r = r_0$. We assume that this value of r corresponds to point A or B in Figure 5, that is, to the "perihelion". At these points r ceases to decrease, that is, achieves the minimum, and $\dot{r} = 0$. Point $r = r_0$ is determined from Eq. (5.7):

$$E = \frac{M^2}{2mr_0^2} + U(r_0) \tag{5.11}$$

Kepler's Problem. Equation (5.10) offers a complete solution of the problem of the motion of a mass point in a central field. The answer in quadrature form involves the initial data. If they are known, integration can be carried out in one way or another. The fact that the integral sometimes cannot be solved in terms of elementary functions is not essential.

But, of course, if the solution is obtained in the form of a known and thoroughly studied function, it is of special interest, because it readily lends itself to graphic investigation.

A simple solution in elementary functions can be found only in a few cases. One of them is that of central forces decreasing inversely as the square of the distance. This is the law governing the forces of Newtonian attraction between mass points (or bodies possessing spherical symmetry).

The problem on the motion of two such bodies is known as *Kepler's problem*, since it was Kepler who empirically established the laws for this case from available data on the motions of the planets across the sky. Newton later developed Kepler's laws theoretically from the equations of mechanics and the law of gravitation as a supplementary hypothesis regarding forces of interaction. From thence the systematic development of exact natural science began.

Today the term "Kepler's problem" is applied to any forces inversely proportional to the square of the distance between two moving points, regardless of their nature or sign. Thus, the Coulomb interaction also falls within the scope of Kepler's problem. We assume the constant in the force law $F = a/r^2$ and in $U(r) = a/r$ to be

negative or positive, depending upon whether the investigated particles are mutually attracted or repulsed.

If we replace $M/(mr)$ in (5.10) by a new variable x, the integral in Kepler's problem is reduced to the form

$$\varphi = - \int_{x_0}^{x} \left(-x^2 - \frac{2a}{M} x + \frac{2E}{m} \right)^{-1/2} dx$$

$$= \text{arc cos} \ \frac{x + a/M}{(a^2/M^2 + 2E/m)^{1/2}} \Big|_{x_0 = M/(mr_0)}^{x = M/(mr)}$$

The angle is measured counterclockwise from the perihelion (see Figure 6 on p. 64).

Substitution of the lower integration limit yields zero. This is seen both from (5.10) according to the choice of the origin for φ, and directly from the equation. Indeed, when the radicand vanishes, we have unity under the arc cos sign, and arc cos $1 = 0$.

Inverting the integration result and turning to the variable r, we obtain after some simple transformations

$$r = \frac{M^2}{am} \left[-1 + \frac{M}{a} \left(\frac{a^2}{M^2} + \frac{2E}{m} \right)^{1/2} \cos \varphi \right]^{-1} \qquad (5.12a)$$

This formula is valid for both signs of the force constant a. Let us write it for each sign separately. First let $a > 0$, which corresponds to repulsive forces. Then, after a slight transformation of (5.12a), we obtain for the path of a point in a force field the equation of a hyperbola

$$r = \frac{M^2}{am} \left[\left(1 + \frac{2EM^2}{ma^2} \right)^{1/2} \cos \varphi - 1 \right]^{-1} \qquad (5.12b)$$

Its eccentricity is $[1 + 2EM^2/(ma^2)]^{1/2}$ and consequently greater than unity. At $\varphi = 0$ the denominator of the fraction has its greatest value, and r its smallest (the perihelion). But as φ increases cos φ decreases, and at some point the denominator vanishes, while r tends to infinity. The corresponding angle φ_0 gives the direction of the asymptote to the hyperbola (see Figure 6). Greater values of φ are meaningless since the radius vector of the mass point corresponding to them would be negative.

For forces of attraction, $a < 0$, and from (5.12) we have

$$r = \frac{M^2}{|a|m} \left[\left(1 + \frac{2EM^2}{ma^2} \right)^{1/2} \cos \varphi + 1 \right]^{-1} \qquad (5.12c)$$

Now the path may be of two forms, depending on the sign of the energy. At $E > 0$ the eccentricity of the curve is again greater than unity, and we have a hyperbolic trajectory. But since we now have a plus sign in the denominator the angle φ at which the path extends

into infinity is greater than $\pi/2$. This means that a particle approaching from infinity is deflected from the centre of attraction and skirts it. It encloses the focus of the curve within the asymptotes, while the hyperbola in Figure 6, drawn for the case of repulsive forces, corresponds to the focus being outside the angle between the asymptotes. This circumstance is quite obvious.

If $E < 0$, the eccentricity is less than unity. Since $\cos \varphi$ is also always smaller than unity, the denominator in (5.12c) never vanishes, and we obtain the equation of an ellipse; $\varphi = 0$ corresponds to the perihelion, and $\varphi = \pi$ to the aphelion.

EXERCISES

1. A mass point m is travelling towards an attraction centre for which the potential energy expression is $-\mid a \mid/r^2$. The velocity of the particle at infinity is given in magnitude and direction. A straight line is drawn from the centre parallel to that direction. The distance between the line and the path at infinity is ρ. Determine the value of ρ at which the paths receding again into infinity and the paths spiraling towards the centre separate.

2. Obtain the path equation for the case of $U = \mid a \mid r^2/2$, $E > 0$.

Answer. A circle or an ellipse; unlike the case of Kepler's problem, the centre of attraction lies at the centre of the orbit.

3. Prove that Kepler's problem involves a supplementary integral of the motion expressed in the form of the vector

$$\mathbf{N} = \mathbf{v} \times \mathbf{M} - \frac{a}{r}\, \mathbf{r}$$

Solution. Differentiate $\mathbf{N}$ with respect to time and then substitute $m\mathbf{r} \times \mathbf{v}$ for $\mathbf{M}$, and $-a\mathbf{r}/r^3$ for $m\dot{\mathbf{v}}$ to get

$$\dot{\mathbf{N}} = \dot{\mathbf{v}} \times \mathbf{M} - \frac{a}{r}\, \mathbf{v} + \frac{a}{r^3}\, \mathbf{r}\,(\mathbf{r} \cdot \mathbf{v})$$

$$= -\frac{a}{r^3}\, \mathbf{r} \times (\mathbf{r} \times \mathbf{v}) - \frac{a}{r^3}\,[r^2\mathbf{v} - \mathbf{r}\,(\mathbf{r} \cdot \mathbf{v})] = 0$$

Vector $\mathbf{N}$ is directed from the focus to the perihelion and is numerically equal to $\mid a \mid e$, where e is the eccentricity of the orbit, $e = [1 - 2\mid E \mid \times M^2/(ma^2)]^{1/2}$.

The fact that vector $\mathbf{N}$ is constant is closely connected with the form of the force law in Kepler's problem, where the path has the form of an ellipse fixed in space. For any other dependence of the potential energy on distance (except for the case given in Exercise 2) the integral (5.10) computed between the two positions of minimum approach of the attracting particles

is not a simple multiple of 2π (or of π, as in Exercise 2). At least, arbitrary values of the integrals E and M do not yield a simple multiple. But this means that in this case the path does not have the form of a closed curve, that is, the "perihelion" rotates in space. Accordingly, the path takes the form of a "rosette". The result is also a rosette when Kepler's problem is solved according to the laws deriving from relativity theory (see Exercise 9, Section 14).

6

COLLISION OF PARTICLES

The Significance of Collision Problems. In order to determine the forces acting between particles, it is necessary to study the motion of the particles caused by these forces. In this way Newton's gravitational law was established with the aid of Kepler's laws. Here, the forces were determined from finite motion. However, infinite motion can also be used if one particle can in some way be accelerated to a definite velocity and then made to pass close to another particle. Such a process is termed "collision" of particles. It is not at all assumed, however, that the particles actually come into contact in the sense of "collision" in everyday life.

Neither is it necessary that the incident particle should be artificially accelerated in a machine: it may be obtained in ejection from a radioactive nucleus or as the result of a nuclear reaction, or it may be a fast particle in cosmic radiation.

Two approaches are possible to problems on particle collisions.

Firstly, it may be only the velocities of the particles long before the collision (before they begin to interact) that are given, and the problem is to determine their velocities (magnitude and direction) after they have ceased to interact, that is, when they have receded to an infinite distance. To solve this kind of problem it is necessary to state some quantities characterizing the collision, for example, the change in energy of the colliding particle or its angle of deflection. Then all the other quantities can be determined with the help of conservation laws. Thus, it is the result of the collision that is summed up, without going into the details of its course.

However, another approach is possible: it is required to precalculate the final state where the precise initial state is given.

We shall first consider collisions by the first method. It is clear why, if only the initial velocities of the particles are known, the collision is not completely determined: it is not known at what distance the particles pass by each other, since we do not know their

initial positions. Therefore some quantity relating to the final state of the system must be given. Usually the problem is stated as follows: the initial velocities of the colliding particles and also the direction of velocity of one of them after the collision are specified. It is required to determine all the remaining quantities after the collision.

If the total kinetic energy of the particles is the same before and after the collision or if its change as a result of the collision is stated exactly, the problem has one solution. There are six unknown quantities: the six angular momentum components of both particles. The conservation laws give four equations: one corresponding to the conservation of the energy (taking into account possible dissipation if the collision is inelastic), and three expressing the conservation of the vector quantity of total linear momentum.

A collision is *inelastic* when a portion of the kinetic energy of the colliding particles is transferred to the internal degrees of freedom. In that case the energy balance must include the portion "trapped" by the internal degrees of freedom, and it must be stated in advance. If the kinetic energy of the colliding particles does not change, the collision is called *elastic*.

One of the quantities characterizing the state of the particles after the collision is usually of no interest: the plane in which the linear momenta of the receding particles lie. They can be arbitrarily stated to be flying apart in, for example, the plane of a diagram in which both final linear momenta are depicted. Thus for the six required quantities we have four equations and one arbitrary plane. It is therefore necessary to state one more quantity characterizing the collision, for instance, the angle of deflection of the incident particle.

There exists a more general type of inelastic collisions in which not only the internal energy of the particles changes but their nature as well. That is what occurs in nuclear reactions. It is then necessary to state the masses of the resultant particles. But strictly speaking this case cannot be treated in the framework of Newtonian mechanics, since it is necessary to take into account the equivalence of mass and energy in accordance with the mechanics of Einstein's relativity theory (see Part II). We shall consider such collisions for the case when the change in total mass can be neglected.

The Laboratory and Centre-of-Mass Frames of Reference. When collisions are observed in laboratory conditions, one of the particles— belonging to the target—is usually at rest prior to the collision. A frame of reference fixed with respect to the target (or the laboratory) is called the *laboratory frame of reference*. In it the colliding particles have a total linear momentum equal to the momentum of the incident particle; the linear momentum of the second particle in this system is, by definition, zero. In accordance with the law of

conservation of linear momentum, the scattering particles must possess the same total momentum.

It follows from this that both these particles also possess the kinetic energy of the common motion of their centre of mass, which is given by formula (3.17); namely, the first term in the right-hand side. Thus, part of the kinetic energy of the incident particle prior to the collision must necessarily be imparted to the centre of mass of the particles following the collision. This portion of the energy cannot be usefully expended if a transmutation act requiring an expenditure of energy results from the collision. That is why it is common to employ a frame of reference fixed with respect to the centre of mass of the particles.

In this system the total linear momentum of the colliding particles is zero; initially they are moving towards one another, after the collision they are receding in strictly opposite directions; in the most general case, at some angle to the initial direction of the linear momenta. In the centre-of-mass reference frame the total energy of the particles prior to the collision may be expended on a transmutation. Obviously, in this case the collision is inelastic to the highest degree.

Differentiation of formula (3.12) with respect to time yields the velocity of the centre-of-mass frame relative to the laboratory frame:

$$\mathbf{V} = \frac{m_1}{m_1 + m_2}\, \mathbf{v}_0 \tag{6.1}$$

Here, $\mathbf{v}_0$ is the velocity of the first particle relative to the second, m_1 is the mass of the first particle, m_2 is the mass of the second particle, which was at rest prior to the collision, $\mathbf{V}$ is the required velocity of the centre-of-mass frame relative to the laboratory frame. Apparently, the quantity $\mathbf{v}_0$, that is, the relative velocity of the particles, is the same in both reference frames. This holds as long as the simple law of velocity composition can be applied, that is, as long as the velocities are small in comparison with the speed of light.

The General Case of an Inelastic Collision. The velocity of the first particle relative to the centre-of-mass frame is, according to the law of addition of velocities,

$$\mathbf{v}_{10} = \mathbf{v}_0 - \mathbf{V} = \frac{m_2}{m_1 + m_2}\, \mathbf{v}_0 \tag{6.2}$$

and in the same reference frame, the velocity of the second particle is

$$\mathbf{v}_{20} = -\mathbf{V} = -\frac{m_1}{m_1 + m_2}\, \mathbf{v}_0 \tag{6.3}$$

Thus, $m_1\mathbf{v}_{10} + m_2\mathbf{v}_{20} = 0$, as it should be in this reference frame.

In accordance with (3.17), the energy in the centre-of-mass frame is

$$E_0 = \frac{m_1 m_2}{2(m_1 + m_2)} v_0^2 \equiv \frac{m_0 v_0^2}{2} \qquad (6.4)$$

Here, the reduced mass is indicated by a zero subscript because in nuclear reactions it may change. Without invoking considerations of mass-energy equivalence, let us simply agree that as a result of the reaction a certain energy Q is released or absorbed, that is, energy transforms from internal energy (corresponding to internal degrees of freedom) into the kinetic energy of the particles $(Q > 0)$ or, conversely, it transforms from kinetic energy into the energy of the internal state $(Q < 0)$.

In the case of nuclear reactions Q is the energy associated with rearrangement in the system. The same is true of chemical reactions, but here it should be noted that two colliding atoms cannot turn into two other atoms, while colliding molecules cannot even be regarded as points (their structure is important). Collisions between neutral atoms may involve changes in their internal energy states, in which case the formulas derived here are also applicable.

Taking into account the energy Q, the energy conservation law in a collision must be written thus:

$$\frac{m_0 v_0^2}{2} + Q = \frac{m v^2}{2} \qquad (6.5)$$

Here, $m = m_3 m_4 / (m_3 + m_4)$ is the reduced mass of the particles produced in the collision, and v is their relative velocity.

In order to specify the collision completely, we consider that the direction of $\mathbf{v}$ is known: its absolute value is given by (6.5). The velocities of each particle separately are

$$\mathbf{v}_{30} = \frac{m_4}{m_3 + m_4} \mathbf{v}, \qquad \mathbf{v}_{40} = -\frac{m_3}{m_3 + m_4} \mathbf{v} \qquad (6.6)$$

They satisfy the requirement $m_3 \mathbf{v}_{30} + m_4 \mathbf{v}_{40} = 0$, that is, the law of conservation of momentum in the centre-of-mass reference frame. It is easy to verify that the energy conservation law also holds, since

$$\frac{m_3 v_{30}^2}{2} + \frac{m_4 v_{40}^2}{2} = \frac{m v^2}{2}$$

Now it is not difficult to revert to the laboratory frame of reference. The velocities of the particles in this frame are

$$\mathbf{v}_3 = \mathbf{v}_{30} + \mathbf{V} = \frac{m_4 \mathbf{v}}{m_3 + m_4} + \frac{m_1 \mathbf{v}_0}{m_1 + m_2}$$

$$\mathbf{v}_4 = \mathbf{v}_{40} + \mathbf{V} = -\frac{m_3 \mathbf{v}}{m_3 + m_4} + \frac{m_1 \mathbf{v}_0}{m_1 + m_2} \qquad (6.7)$$

In nuclear reactions the change in the total mass of the particles is no more than a fraction of a percentage point, so that in Eqs. (6.7) $m_1 + m_2$ can be substituted for $m_3 + m_4$. Thus, knowing $\mathbf{v}/v$, we obtain the complete solution of the problem. Usually the direction of the emerging particles depends upon the position of the detecting device relative to the target. If the energy is measured at the time of recording, the *energy effect* Q of the reaction can be computed.

Elastic Collisions. The computations are simplified if the collision is elastic, for then $m_3 = m_1$, $m_4 = m_2$, $Q = 0$. It follows from (6.5) that $v_0 = v$, so that the relative velocity changes only in direction and not in magnitude. Let us suppose that its angle of deflection χ is given. We take the x axis along $\mathbf{v}_0$ in the plane determined by vectors $\mathbf{v}_0$ and $\mathbf{v}$. Then

$$v_x = v_0 \cos \chi, \qquad v_y = v_0 \sin \chi$$

From (6.6) the components of the particles' velocity in the laboratory reference frame will, after collision, be correspondingly equal to

$$v_{1x} = \frac{(m_1 + m_2 \cos \chi)\, v_0}{m_1 + m_2}, \qquad v_{1y} = v_{10y} = \frac{m_2 v_0 \sin \chi}{m_1 + m_2} \qquad (6.8)$$

$$v_{2x} = \frac{m_1 (1 - \cos \chi)\, v_0}{m_1 + m_2}, \qquad v_{2y} = v_{20y} = -\frac{m_1 v_0 \sin \chi}{m_1 + m_2} \qquad (6.9)$$

By means of these equations, the deflection angle Θ of the first particle in the laboratory frame of reference can be related to the deflection angle χ in the centre-of-mass reference frame in the following manner:

$$\tan \Theta = \frac{v_{1y}}{v_{1x}} = \frac{m_2 \sin \chi}{m_1 + m_2 \cos \chi} \qquad (6.10)$$

The "recoil" angle of the second particle, which before collision was fixed in the laboratory frame of reference, is given by the formula

$$\tan \Theta' = -\frac{v_{2y}}{v_{2x}} = \frac{\sin \chi}{1 - \cos \chi} = \cot \frac{\chi}{2}, \qquad \Theta' = \frac{\pi - \chi}{2} \quad (6.11)$$

Angle Θ' is taken with respect to the x axis; the minus in the definition of $\tan \Theta'$ is chosen because the signs of v_{1y} and v_{2y} are opposite.

Equation (6.10) becomes still simpler if the masses of the colliding particles are equal. This is true if the two particles are, say, protons, and is approximately true if one is a proton and the other a neutron. Then, from (6.10),

$$\tan \Theta = \tan \frac{\chi}{2}, \qquad \Theta = \frac{\chi}{2}$$

$$\Theta' = \frac{\pi - \chi}{2}, \qquad \Theta + \Theta' = \frac{\pi}{2} \qquad (6.12)$$

so that the particles fly apart at right angles, while the deflection angle of a neutron in the laboratory frame is half its deflection angle in the centre-of-mass frame. Since the latter varies from 0 to 180°, Θ never exceeds 90°. This is obvious without computations, because the masses of the particles are equal. In a head-on collision the incident particle comes to a halt while the resting particle is propelled straight ahead with the speed of the incident particle. This is because in a head-on collision in the centre-of-mass frame particles of equal mass simply exchange velocities. In the laboratory frame the second particle thus receives the initial velocity of the first, which comes to a halt, as was asserted.

We shall now determine the kinetic energy transferred to the second particle in the collision, starting with the case of different masses. From Eqs. (6.9) we find

$$E_2 = \frac{m_2}{2}\left(v_{2x}^2 + v_{2y}^2\right) = \frac{m_2 m_1^2 \left(1 - \cos\chi\right) v_0^2}{(m_1 + m_2)^2}$$

Relative to the energy E_0 of the incident particle, this is

$$\frac{E_2}{E_0} = \frac{2m_1 \left(1 - \cos\chi\right)}{m_1 + m_2} \tag{6.13}$$

For particles of equal mass we obtain

$$\frac{E_2}{E_0} = \sin^2\frac{\chi}{2} = \sin^2\Theta$$

Accordingly, for the first particle there remains

$$\frac{E_1}{E_0} = \cos^2\Theta$$

In a head-on collision, $\chi = 180°$, $\Theta = 90°$, $E_1 = 0$, and $E_2 = E_0$, as was proved.

The Particle Scattering Problem. We shall now consider the collision problem in greater detail, confining ourselves to the case of elastic collision, which is more conveniently investigated in the centre-of-mass frame of reference. The transformation to the laboratory frame according to Eqs. (6.7) is straightforward.

Obviously, for a complete solution of the collision problem we must know the potential energy $U(r)$ of the particles' interaction and state the initial conditions in a way that would make it possible to determine all the integrals of the motion. The energy integral is easily found if we recall that at infinity U is gauged to zero: $U(\infty) = 0$. Denoting the relative velocity of particles at an infinite distance from each other by v, we obtain the value of the energy integral:

$$E = \frac{mv^2}{2} \tag{6.14}$$

Now we determine the angular momentum integral. Figure 6 presents the motion of the first particle relative to the second for the case of repulsive forces. Its path at infinity from the second particle is a straight line, because the forces vanish there. Hence, the path of the particle has asymptotes both along the approach section (line AF) and along the receding section (line FB). The distance ρ of the asymptote AF from line OC drawn parallel to it from the second
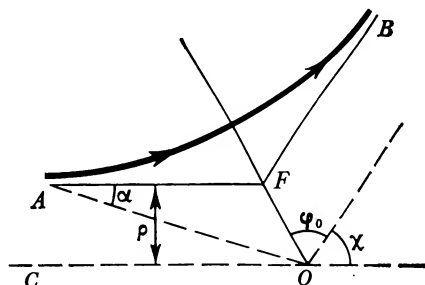

Figure 6

particle is known as the *impact*, or *collision, parameter*; it is nothing other than the "arm" when the particles are infinitely apart. It is apparent from this that the angular momentum is

$$M = mv\rho \tag{6.15}$$

The mass in Eqs. (6.14) and (6.15) is the reduced mass.

Knowing the energy and angular momentum integrals we can compute the deflection angle of the first particle. From Figure 6 it can be seen that this angle is connected with the angle $2\varphi_0$ between the asymptotes by the simple relationship $\chi = \pi - 2\varphi_0$. In turn, φ_0 is calculated with the help of the quadrature according to Eq. (5.10), where the upper limit is taken equal to infinity:

$$\varphi_0 = \int\limits_{r_0}^{\infty} \frac{v\rho}{r^2} \frac{dr}{[v^2 - v^2\rho^2/r^2 - 2U\,(r)/m]^{1/2}} \tag{6.16}$$

The angular momentum and energy integrals are already substituted here using (6.14) and (6.15). The lower limit is calculated from Eq. (5.11).

The Differential Cross Section. Suppose the integral in (6.16) can be found. Then φ_0 and hence the deflection angle χ are known as functions of the impact parameter. Let this function be inverted, that is, the impact parameter is obtained as a function of the de-

flection angle:

$$\rho = \rho(\chi) \tag{6.17}$$

In collision experiments the impact parameter is actually never known in advance: a parallel beam of scattering particles having the same velocity is directed on a substance whose atoms or nuclei are the scatterers. The distribution of particles according to their
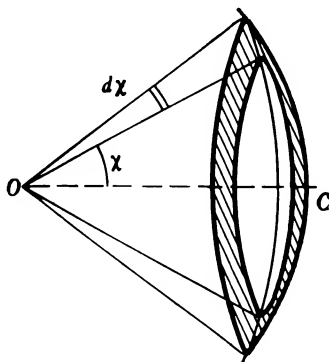

Figure 7

deflection angles χ, more precisely, according to the deflection angles Θ in the laboratory system, is observed. Thus, the scattering events are, so to say, observed at many successive times, with the most diverse impact parameters.

Let one particle pass through one square centimetre of the scattering substance. Then $2\pi\rho \, d\rho$ particles pass through an annulus contained between ρ and $\rho + d\rho$. We thus classify the collisions according to impact parameters much as is done on a shooting range with the help of targets with a pattern of concentric rings. If ρ is known as a function of χ, it can be asserted that $d\sigma \equiv 2\pi\rho \, d\rho = 2\pi\rho \, (d\rho/d\chi) \, d\chi$ particles will be deflected at an angle lying between χ and $\chi + d\chi$. The dependence of ρ on χ is determined from Eq. (6.17).

Suppose the scattered particles are in some way detected at a large distance from the scattering medium. The latter can then be treated as a point and the scattered particles assumed to be travelling along rectilinear paths radiating from a common centre. Let us examine the particles moving in the space between two cones sharing a common apex and common axis. The direction of the axis coincides with the direction of the incident particles. The half-angle at the apex of the inner cone is equal to χ, and the half-angle of the outer cone is $\chi + d\chi$ (Figure 7).

The space between the two cones is called a solid angle, by analogy with a plane angle, which is defined as the part of a plane between two straight lines. The measure of a plane angle is the arc of a circle of unit radius with its centre at the apex of the angle; the measure of a solid angle is the area of the sector of a sphere of unit radius drawn from the apex and contained within the cone.

In Figure 7 the solid-angle element is represented as part of the surface of a sphere swept out by the arc element $d\chi$ in a rotation about the radius OC. Since $OC = 1$, the radius of rotation of the element $d\chi$ is equal to $\sin \chi$, whence the area of the sphere swept out by it is $2\pi \sin \chi \, d\chi$. Thus, the solid-angle element is

$$d\Omega = 2\pi \sin \chi \, d\chi \qquad (6.18)$$

In passing from the angle element $d\chi$ to the elementary solid-angle element $d\Omega$ we can write the expression for the number of particles scattered within the solid-angle element:

$$d\sigma = \rho \, \frac{d\rho}{d\chi} \, \frac{d\Omega}{\sin \chi} \qquad (6.19)$$

The quantity $d\sigma$ has the dimensions of area. It is the area within which a particle must fall to be scattered within the elementary solid-angle element $d\Omega$. It is called the *differential cross section* of scattering into the solid-angle element $d\Omega$.

In experiments it is this quantity that is determined in recording the particles deflected at different angles. If a unit volume of the scattering medium contains n scatterers, the attenuation of the primary parallel beam in passing through unit thickness of the medium due to deflection into the *solid-angle element $d\Omega$* is

$$dJ = -Jn \, d\sigma = -Jn\rho \, \frac{d\rho}{d\chi} \, \frac{d\Omega}{\sin \chi} \text{ particles/(cm-s)}$$

In studying the dependence of $d\sigma$ on χ we find how the impact parameter depends on the deflection angle, which makes it possible to draw conclusions about the nature of the forces acting between the particle and the scattering centre.

Rutherford's Formula. The most important application of Eq. (6.19) is in particle scattering in a Coulomb field. Assuming the scattering medium to have very large mass in comparison with the scattered charged particle, the laboratory frame of reference differs but slightly from the centre-of-mass frame, and the reduced mass closely approximates the mass of the light participant in the collision (see 3.21)). As was pointed out in Section 3, the Coulomb potential decreases with distance according to the law $1/r$, like the Newtonian gravity potential. Consequently, the deflection angle can be computed from Eqs. (5.12b) and (5.12c), depending on whether the forces acting between the particles are of attraction or repulsion.

Let the charge of the scatterer be *Ze*, and that of the scattered particle $\pm e$. Then the force constant *a* is equal to $\pm Ze^2$. The colliding particles are at infinite distance when the denominators in Eqs. (5.12b) and (5.12c) vanish, that is, when

$$\cos \varphi_0 = \pm \left(1 + \frac{2M^2E}{ma^2} \right)^{-1/2}, \quad \text{or} \quad \tan \varphi_0 = \left(\frac{2M^2E}{ma^2} \right)^{1/2} \tag{6.20}$$

Since only the absolute value of the deflection angle matters in determining the differential cross section, we take one sign for the tangent. Different signs of *a* in Figure 6 would correspond to deflections of the particle up or down from the page, which in this case is irrelevant.

Determining the integrals of the motion according to Eqs. (6.14) and (6.15), and recalling that $\chi = \pi - 2\varphi_0$, we find the impact parameter as a function of the deflection angle:

$$\rho = \frac{a}{mv^2} \cot \frac{\chi}{2}$$

Since in the present case the centre-of-mass frame of reference is closely approximated by the laboratory frame, χ can be replaced by Θ, that is, the particle's deflection angle in the laboratory frame. We now write the formula of the differential cross section in the laboratory frame according to the general definition (6.19):

$$d\sigma = \frac{Z^2e^4}{4m^2v^4} \frac{d\Omega}{\sin^4(\Theta/2)} \tag{6.21}$$

The number of particles scattered into the solid-angle element $d\Omega = 2\pi \sin \Theta \, d\Theta$ is inversely proportional to the fourth power of the sine of one-half the deflection angle. This law is uniquely connected with the Coulomb character of the scattering forces. The greater the deflection angle the less the impact parameter.

E. Rutherford traced the law (6.21) for the case of scattering of alpha particles on heavy nuclei up to very small impact parameters. At impact parameters of the order 8×10^{-13} cm, scattering becomes subject to a different law. Rutherford concluded from this that the whole positive charge of an atom is concentrated at the centre, since the diameter of an atom is around 10^{-8} cm. Practically the whole mass of the atom is also concentrated at the centre, otherwise alpha particles could not be scattered at large angles (sometimes their initial directions are almost reversed). Thus Rutherford's experiments with alpha scattering led to the discovery of the atomic nucleus.

If the colliding particles are of nearly the same mass, in Eq. (6.21) the deflection angle χ in the centre-of-mass reference frame should be substituted for angle Θ, and the mass of the lighter participant

substituted for the reduced mass, after which one can go over to the laboratory reference frame according to (6.10) and (6.11).

The Rutherford formula undergoes a curious change in collisions of two identical particles, for example, in the scattering of an alpha particle on a helium nucleus (the particles are, as is known, the same). There is no method capable of distinguishing one alpha particle from another. It is impossible to determine whether a detected particle was at rest prior to the collision or impinged on a resting particle. Therefore, to obtain a formula suitable for comparison with experiment it is necessary to take all the particles into account in the expression for $d\sigma$.

It follows from Eqs. (6.12) that to transfer from the centre-of-mass frame to the laboratory frame in collisions of identical particles, χ must be replaced by 2Θ. Then $\sin \chi \, d\chi = 2 \sin 2\Theta \, d\Theta$. It should also be taken into account that in the laboratory frame the particles scatter at right angles, so that $\sin \Theta' = \cos \Theta$. Substituting the reduced mass $m^2/(2m) = m/2$ for the actual mass m, we obtain the differential cross section for two identical particles interacting according to the Coulomb law:

$$d\sigma = \frac{4\pi Z^2 e^4}{m^2 v^4} \left(\frac{1}{\sin^4 \Theta} + \frac{1}{\cos^4 \Theta} \right) \sin 2\Theta \, d\Theta \qquad (6.22)$$

Isotropic Scattering. From Eq. (6.21) it is apparent that scattering has a pronounced maximum for small deflection angles. This maximum is associated with large impact parameters: particles passing each other at great distances deflect weakly, and great impact parameters predominate in the expression for the differential cross section, since they involve a greater area. Therefore, if the forces of interaction between the particles do not identically vanish at finite distances, any particle, however far from the scatterer it may pass, is somewhat deflected. The ratio $d\sigma/d\Omega$ then inevitably tends to infinity at small deflection angles.

But if at large distances the force is not exactly zero but closely approaches it, that is, decreases rapidly, then the differential cross section begins to increase appreciably, tending to infinity only at the smallest deflection angles. But in experiments weakly deflected particles are not detected at all as having been deflected. Indeed, the initial beam already possesses a certain scattering, and deflection angles lying within this scattering angle cannot be detected.

If the force decreases very rapidly with distance, the domain of the sharply increasing $d\sigma/d\Omega$ as a function of angle χ may fall within such small angles that they cannot be experimentally detected as deflection angles, that is, separated from the initial beam. On the other hand, all strongly deflected particles are the more uniformly distributed as regards scattering angles the faster the forces diminish with distance.

This is shown in the example of particle scattering by an impermeable sphere (Exercise 1). The acting force can be treated as the limiting case of a force centre repulsing particles according to the law $U(r) = U_0 (r_0/r)^n$ at n tending to infinity. If $r < r_0$, then $U(r) \to \infty$; and if $r > r_0$, then $U(r) \to 0$. In other words, at $n \to \infty$ particles cannot penetrate the domain where $r > r_0$, which corresponds to an impermeable sphere. But as can be seen from Exercise 1 at $n = \infty$ the scattering is completely isotropic. If n is not infinite but sufficiently large, the particles are distributed almost isotropically over all angles, and a sharp maximum appears only at small deflection angles, receding into infinity when the deflection angle tends to zero. Consequently, a scattering law approximating the isotropic law is indicative of a rapid diminishing of the forces with distance. This circumstance played an important part in the investigation of nuclear forces.

EXERCISES

1. Find the differential cross section for particles scattering on an impermeable sphere of radius r_0.

Solution. An impermeable sphere can be described in terms of mechanics by stating the potential energy in the form $U(r) = 0$ at $r > r_0$ (outside the sphere) and $U(r) = \infty$ at $r \leqslant r_0$ (inside the sphere). Then whatever a particle's kinetic energy, its penetration into the domain $r < r_0$ is impossible.

Reflection from the sphere takes place in the following way. The radial momentum component reverses its sign, while the tangential component is conserved, since, given radial symmetry of the potential, no forces can be perpendicular to the radius. The absolute value of the linear momentum is conserved, insofar as the impact is elastic and the kinetic energy does not change. A simple construction reveals that the impact parameter is related to the deflection angle by the dependence $\rho = r_0 \cos(\chi/2)$ if $\rho < r_0$. Hence the general formula yields

$$d\sigma = (r_0^2/4)\, d\Omega$$

so that the scattering is uniform over all angles, that is, it is isotropic. The total scattering cross section σ is equal to πr_0^2 in this case, as could be expected. Significant here is the fact that the interaction forces identically vanish at a finite distance.

2. A collision of two particles is observed, m_1 and m_2 being their masses (m_1 is the mass of the incident particle). As a result of the collision particles are formed whose angular momenta lie at angles φ and ψ to the angular momentum of the incident particle. Determine the energy Q by which the total kinetic energy of the colliding particles changes. Consider two cases: the collision yields (a) two particles of the same masses m_1 and m_2 and (b) two particles of different masses m_3 and m_4, in sum equal to $m_1 + m_2$.

7

SMALL OSCILLATIONS

In Section 4 we considered the oscillations of a pendulum. It was pointed out that they obey a complex law, which cannot be described with the help of elementary functions. The situation is simplified, however, in the limiting case of very small deflection angles of the pendulum from the vertical. This type of motion, known as *small oscillations*, is very important in applications of mechanics and, up to various approximations, is widely encountered in nature and in technology. For this reason the theory of small oscillations is treated here as a special section.

Small Oscillations of a Pendulum. A simple graphic investigation reveals that in the most general case the motion of a pendulum is periodic. Figure 8 shows the curve $U(\varphi) = mgl(1 - \cos \varphi)$, which
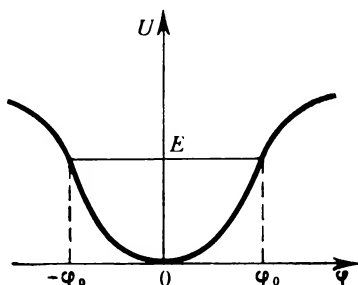

Figure 8

gives the relationship between potential energy and deflection angle. The horizontal straight line corresponds to a constant value of E. If $E < 2mgl$, the motion occurs periodically with time between the points $-\varphi_0$ and φ_0.

The problem is greatly simplified if $\varphi_0 \ll 1$, that is, the angle is small in comparison with one radian. Then $\cos \varphi_0$ can be replaced by a Taylor expansion up to the second term: $\cos \varphi_0 = 1 - \varphi_0^2/2$. Since $|\varphi| < \varphi_0$, the expansion is also valid for $\cos \varphi$. After this the integral (4.7) can be easily evaluated:

$$t = -\left(\frac{l}{g}\right)^{1/2} \int_{\varphi_0}^{\varphi} \frac{d\psi}{(\varphi_0^2 - \varphi^2)^{1/2}} = \left(\frac{l}{g}\right)^{1/2} \text{arc} \cos \frac{\varphi}{\varphi_0} \qquad (7.1)$$

Inverting relation (7.1), we get the deflection angle as a function of time:

$$\varphi = \varphi_0 \cos t(g/l)^{1/2} \tag{7.2}$$

The result is a periodic function. As can be seen from (7.2), the deflection angle reverts to its initial value in time $\tau = 2\pi(l/g)^{1/2}$, which is known as the *period of oscillation*. The quantity $(g/l)^{1/2}$ is called the *frequency of oscillation*:

$$\omega \equiv (g/l)^{1/2} \tag{7.3}$$

This quantity gives the number of radians by which the argument of the cosine in (7.2) changes in one second. The term *frequency* is also used to denote the number of periods per second; in that case it is smaller than ω by a factor of 2π. The oscillation period τ is connected with ω by the relationship $\tau = 2\pi/\omega$. It is significant that the period and frequency of small oscillations do not depend on their amplitude φ_0.

The General Problem of Small Oscillations with One Degree of Freedom. In solving the problem of small oscillations there is no need to first reduce to quadrature the problem of arbitrary oscillations. We can initially simplify the Lagrangian.

First, we note that all oscillations, large and small, always occur around a position of equilibrium (for instance, a pendulum oscillates about its vertical position). When deflected from stable equilibrium a system is subject to a force which acts in the opposite direction of the deflection, whatever its sign. This is known as the *restoring force*. At the equilibrium point the force is obviously zero, simply by definition of the concept of equilibrium.

Force is equal to the derivative of potential energy taken with the opposite sign. We shall take the derivative with respect to the generalized coordinate. Then the equilibrium condition has the form

$$\frac{\partial U}{\partial q} = 0 \tag{7.4}$$

Let us denote the solution of this equation $q = q_0$. Assuming the system to have only one degree of freedom, we expand $U(q)$ in a Taylor series in the vicinity of point q_0 up to the quadratic term to get

$$U(q) = U(q_0) + \left(\frac{\partial U}{\partial q}\right)_{q=q_0}(q - q_0)$$
$$+ \frac{1}{2}\left(\frac{\partial^2 U}{\partial q^2}\right)_{q=q_0}(q - q_0)^2 + \ldots \tag{7.5}$$

In accordance with (7.4), the term linear in $(q - q_0)$ vanishes. We denote $(\partial^2 U/\partial q^2)_{q=q_0}$ by β. Then, confining ourselves to the

indicated terms of the series, we obtain

$$U(q) \approx U(q_0) + \frac{\beta}{2}(q - q_0)^2 \tag{7.6}$$

The force near the equilibrium position is

$$F(q) = -\frac{\partial U}{\partial q} = -\beta(q - q_0) \tag{7.7}$$

For this force to be a restoring force, that is, for the equilibrium to be stable, the following inequality must hold:

$$\beta = \left(\frac{\partial^2 U}{\partial q^2}\right)_{q=q_0} > 0 \tag{7.8}$$

This is the stability condition for the equilibrium: the function $U(q)$ must increase on both sides of the point $q = q_0$. It follows that the potential energy at that point must have a minimum. This is shown in Figure 8 at $\varphi = 0$.

Let us now examine the expression for kinetic energy. If in the general kinetic energy formula we substitute $x = x(q)$, $y = y(q)$, and $z = z(q)$, then T reduces to the form

$$T = \frac{m}{2}(\dot{x}^2 + \dot{y}^2 + \dot{z}^2) = \frac{m}{2}\left[\left(\frac{dx}{dq}\right)^2 + \left(\frac{dy}{dq}\right)^2 + \left(\frac{dz}{dq}\right)^2\right]\dot{q}^2$$

The quantity in the brackets depends only on q; and so the kinetic energy of a particle can be represented in the form

$$T = \frac{1}{2}\alpha(q)\dot{q}^2 \tag{7.9}$$

Let us now expand the function $\alpha(q)$ in a series of $q - q_0$ in the vicinity of the equilibrium position:

$$T = \frac{1}{2}\alpha(q_0)\dot{q}^2 + \frac{1}{2}\left(\frac{\partial \alpha}{\partial q}\right)_{q=q_0}(q - q_0)\dot{q}^2 + \ldots$$

In order that the particle should not move far from the equilibrium position, its velocity must be small. In other words, the member of the kinetic energy expansion $\alpha(q_0)\dot{q}^2/2$ is already of the same order for small oscillations as the third term in the expansion of U, that is, $\beta(q - q_0)^2/2$. When $q = q_0$, all the energy of oscillation is kinetic, while at maximum deflection all the energy is potential. But since the total energy is the same at all points, we must assume that the potential and kinetic energies are of the same order of magnitude; this, as always in evaluations, refers to the leading terms in the expansion of T and U. Consequently, in the first approximation the subsequent terms can be neglected, if they are not of some special interest. Later (after the explicit dependence of q on time is

determined from the equation) it will be shown that the mean values of U and T are equal.

In future, the coordinate q will be measured from the equilibrium position, that is, we shall put $q_0 = 0$. Then, omitting $U(0)$ from the Lagrangian, we can write

$$L = \frac{1}{2}\,\alpha\,(0)\,\dot{q}^2 - \frac{1}{2}\,\beta q^2 \tag{7.10}$$

From this we obtain the Lagrange equation

$$\alpha\,(0)\,\ddot{q} + \beta q = 0 \tag{7.11}$$

Denoting

$$\omega^2 = \frac{\beta}{\alpha\,(0)} = \frac{1}{\alpha\,(0)}\left(\frac{\partial^2 U}{\partial q^2}\right)_{q=q_0} \tag{7.12}$$

we reduce (7.11) to the usual form for the oscillation equation:

$$\ddot{q} + \omega^2 q = 0 \tag{7.13}$$

The general solution of this equation must contain two arbitrary constants. It may be written in one of three forms:

$$q = C_1 \cos \omega t + C_2 \sin \omega t \tag{7.14a}$$
$$q = C \cos (\omega t + \gamma) \tag{7.14b}$$
$$q = \mathrm{Re}\,(C' e^{i\omega t}) \tag{7.14c}$$

The symbol Re ( ) signifies the real part of the expression inside the parentheses. The constant C' is complex: $C' = C_1 - iC_2$. The constants C and γ in solution (7.14b) are known as the *amplitude* and the *initial phase* of the oscillation. They are connected with C_1 and C_2 by the known formulas

$$C = (C_1^2 + C_2^2)^{1/2}, \qquad \gamma = -\arctan \frac{C_2}{C_1}$$

If we are only interested in the frequency of small oscillations and not the phase or amplitude, it is sufficient to use Eq. (7.12), verifying that the second derivative $(\partial^2 U/\partial q^2)_{q=q_0}$ is positive.

A system which is described by Eq. (7.13) is called a *linear harmonic oscillator*.

It can be seen from Eqs. (7.10), (7.12), (7.14b) that the averages of the potential energy and kinetic energy during one period are the same, because the averages of the squares of a sine or cosine are

equal to one-half:

$$\overline{\sin^2(\omega t + \gamma)} = \frac{\omega}{2\pi} \int\limits_0^{2\pi/\omega} \sin^2(\omega t + \gamma)\, dt = \frac{1}{2}$$

$$\overline{\cos^2(\omega t + \gamma)} = \frac{1}{2}$$

$$\overline{T} = \overline{U} = \frac{1}{4}\,\alpha(0)\,\omega^2 C^2 = \frac{1}{4}\,\beta C^2$$

Small Oscillations with Two or More Degrees of Freedom. We shall now consider oscillations with two degrees of freedom. As an example, let us first take the double pendulum (Sec. 3). If we confine ourselves to small oscillations, we must consider that the deflections φ and ψ are close to zero (that is, the pendulum oscillates with small deflections close to the vertical). In that case we must substitute the equilibrium values, $\varphi = \psi = 0$ and $\cos(\varphi - \psi) = 1$, into the expression for the kinetic energy in (3.23). The potential energy formula must be simplified in the same way as in the simple pendulum problem, that is, we must replace $\cos\varphi$ and $\cos\psi$ by $1 - \varphi^2/2$ and $1 - \psi^2/2$. Eliminating the constant terms, we obtain the Lagrangian in the form

$$L = \frac{m+m_1}{2}\, l^2\dot\varphi^2 + \frac{m_1}{2}\, l_1^2\dot\psi^2$$
$$+ m_1 l l_1 \dot\varphi\dot\psi - \frac{m+m_1}{2}\, lg\varphi^2 - \frac{m_1 l_1 g}{2}\, \psi^2 \qquad (7.15)$$

Let us examine this in a somewhat more general form:

$$L = \frac{1}{2}\,(\alpha_{11}\dot q_1^2 + 2\alpha_{12}\dot q_1 \dot q_2 + \alpha_{22}\dot q_2^2)$$
$$- \frac{1}{2}\,(\beta_{11} q_1^2 + 2\beta_{12} q_1 q_2 + \beta_{22} q_2^2) - U(0) \qquad (7.16a)$$

Taking into account the summation condition and omitting the constant term, we write

$$L = \frac{1}{2}\,\alpha_{\mu\nu}\dot q_\nu^2 - \frac{1}{2}\,\beta_{\mu\nu} q_\nu^2 \qquad (7.16b)$$

In this form it refers to a system with an arbitrary number of vibrational degrees of freedom.

Comparing (7.15) and (7.16a), we obtain the values of the factors $\alpha_{\mu\nu}$ and $\beta_{\mu\nu}$ for the case of a double pendulum:

$$\alpha_{11} = (m + m_1)\, l^2, \qquad \alpha_{12} = m_1 l l_1, \qquad \alpha_{22} = m_1 l_1^2$$
$$\beta_{11} = (m + m_1)\, lg, \qquad \beta_{12} = 0, \qquad \beta_{22} = m_1 l_1 g$$

The fact that $\beta_{12} = 0$ introduces no simplifications into the problem.

In the most general case the coefficients β_{11}, β_{12}, and β_{22} are expressed by the equations

$$\beta_{11} = \left(\frac{\partial^2 U}{\partial q_1^2}\right)_0, \qquad \beta_{12} = \left(\frac{\partial^2 U}{\partial q_1 \partial q_2}\right)_0, \qquad \beta_{22} = \left(\frac{\partial^2 U}{\partial q_2^2}\right)_0$$

where the equilibrium values are substituted into the derivatives; the former must be determined from equations similar to (7.4):

$$\frac{\partial U}{\partial q_1} = 0, \qquad \frac{\partial U}{\partial q_2} = 0, \qquad \text{or} \qquad \frac{\partial U}{\partial q_\nu} = 0 \qquad (7.17)$$

For the equilibrium to be stable we must require the following inequality to be satisfied:

$$U(q) - U(0) = \frac{1}{2}(\beta_{11}q_1^2 + 2\beta_{12}q_1q_2 + \beta_{22}q_2^2) > 0$$

Given this condition, the minimum of $U(q)$ lies at point $q_1 = 0$, $q_2 = 0$.

Let us rewrite the left-hand side of the inequality in identical form:

$$\frac{1}{2}(\beta_{11}q_1^2 + 2\beta_{12}q_1q_2 + \beta_{22}q_2^2) = \left(q_1 + \frac{\beta_{12}q_2}{\beta_{11}}\right)^2 + \frac{\beta_{22}\beta_{11} - \beta_{12}^2}{4} q_2^2$$

This expression remains positive for all values of q_1 and q_2, provided the coefficients of both quadratics in q are greater than zero:

$$\beta_{11} > 0$$
$$\beta_{11}\beta_{22} - \beta_{12}^2 > 0 \qquad\qquad\qquad\qquad (7.18)$$

The obtained inequalities can be easily generalized for the case of an arbitrary number of degrees of freedom. For this the coefficients must be written in the form of a square array:

$$\beta_{\mu\nu} = \begin{array}{|ccc|} \beta_{11} & \beta_{12} & \beta_{13} & \cdots \\ \beta_{21} & \beta_{22} & \beta_{23} & \cdots \\ \beta_{31} & \beta_{32} & \beta_{33} & \cdots \\ \end{array} \qquad (7.19)$$

Then by the method of induction it can be concluded that all the determinants whose principal diagonal coincides with the principal diagonal of (7.19) must be positive for the quadratic form $\beta_{\mu\nu}q_\mu q_\nu$ to remain greater than zero at all q_μ, q_ν (the proof is presented in any course in higher algebra). In (7.19) these determinants are lined off below and on the right. As for the quadratic form for the kinetic energy, the positive values are automatically assured by the initial expression written in terms of Cartesian coordinates. In future we shall always assume conditions (7.19) to be satisfied.

We shall now write the Lagrange equations:

$$\frac{\partial L}{\partial \dot{q}_1} = \alpha_{11}\dot{q}_1 + \alpha_{12}\dot{q}_2, \qquad \frac{\partial L}{\partial \dot{q}_2} = \alpha_{12}\dot{q}_1 + \alpha_{22}\dot{q}_2, \qquad \frac{\partial L}{\partial \dot{q}_\mu} = \alpha_{\mu\nu}\ddot{q}_\nu$$

$$-\frac{\partial L}{\partial q_1} = \beta_{11}q_1 + \beta_{12}q_2, \quad -\frac{\partial L}{\partial q_2} = \beta_{12}q_1 + \beta_{22}q_2, \quad -\frac{\partial L}{\partial q_\mu} = \beta_{\mu\nu}q_\nu$$

Whence

$$\left.\begin{array}{l} \alpha_{11}\ddot{q}_1 + \alpha_{12}\ddot{q}_2 + \beta_{11}q_1 + \beta_{12}q_2 = 0 \\[4pt] \alpha_{12}\ddot{q}_1 + \alpha_{22}\ddot{q}_2 + \beta_{12}q_1 + \beta_{22}q_2 = 0 \end{array}\right\} \tag{7.20a}$$

or in general form (for any number of degrees of freedom)

$$\alpha_{\mu\nu}\ddot{q}_\nu + \beta_{\mu\nu}q_\nu = 0 \tag{7.20b}$$

In order to satisfy these equations, we shall look for a solution in the form

$$q_1 = A_1 e^{i\omega t}, \qquad q_2 = A_2 e^{i\omega t}, \qquad q_\mu = A_\mu e^{i\omega t} \tag{7.21}$$

As in (7.14c), the real part of the solutions must be taken.

The Frequency Equation. Substituting (7.21) into (7.20), and cancelling out $e^{i\omega t}$, we obtain the equations relating A_1 and A_2:

$$\left.\begin{array}{l} (\beta_{11} - \alpha_{11}\omega^2)\,A_1 + (\beta_{12} - \alpha_{12}\omega^2)\,A_2 = 0 \\[4pt] (\beta_{12} - \alpha_{12}\omega^2)\,A_1 + (\beta_{22} - \alpha_{22}\omega^2)\,A_2 = 0 \end{array}\right\} \tag{7.22a}$$

or in the general case

$$(\beta_{\mu\nu} - \omega^2\alpha_{\mu\nu})\,A_\nu = 0 \tag{7.22b}$$

For a set of linear homogeneous equations to have a solution other than zero the system determinant must vanish:

$$\begin{vmatrix} \beta_{11} - \alpha_{11}\omega^2 & \beta_{12} - \alpha_{12}\omega^2 \\ \beta_{12} - \alpha_{12}\omega^2 & \beta_{22} - \alpha_{22}\omega^2 \end{vmatrix} = 0, \quad \text{or} \quad |\beta_{\mu\nu} - \omega^2\alpha_{\mu\nu}| = 0 \tag{7.23}$$

From this, as applied to (7.22a), we obtain the biquadratic equation

$$(\alpha_{11}\alpha_{22} - \alpha_{12}^2)\,\omega^4 - (\beta_{11}\alpha_{22} + \beta_{22}\alpha_{11} - 2\alpha_{12}\beta_{12})\,\omega^2$$
$$+ \beta_{11}\beta_{22} - \beta_{12}^2 = 0 \tag{7.24}$$

For example, for a double pendulum Eq. (7.24) looks like this:

$$mm_1 l^2 l_1^2 \omega^4 - (m + m_1)\,m_1 l l_1 (l_1 + l)\,g\omega^2 + (m + m_1)m_1 l l_1 g^2 = 0$$

If we introduce the abbreviated notations $l_1/l = \lambda$, $m_1/m = \mu$ specifically for this problem, the expression for frequencies will be of

the following form:

$$\omega^2 = \frac{g}{2l\lambda} \left\{ (1+\mu)(1+\lambda) \pm [(1+\mu)^2(1+\lambda)^2 - 4\lambda(1+\mu)]^{1/2} \right\}$$

It is easy to see that this expression yields only the real values of the frequencies. However, we shall show this in more general form for Eq. (7.24). Let us assume that the following function is given:

$$F(\omega^2) \equiv (\alpha_{11}\alpha_{22} - \alpha_{12}^2)\,\omega^4 - (\beta_{11}\alpha_{22} + \beta_{22}\alpha_{11} - 2\beta_{12}\alpha_{12})\,\omega^2 \\ + \beta_{11}\beta_{22} - \beta_{12}^2$$

which passes through zero at all points where Eq. (7.24) is satisfied. We see that $F(\omega^2)$ is positive for $\omega^2 = 0$ and for $\omega^2 = \infty$, since $\alpha_{11}\alpha_{22} - \alpha_{12}^2 > 0$ and $\beta_{11}\beta_{22} - \beta_{12}^2 > 0$. Let us now substitute into this function the positive number $\omega^2 = \beta_{11}/\alpha_{11}$. After a simple rearrangement we obtain

$$\alpha_{11}^2 F\left(\frac{\beta_{11}}{\alpha_{11}}\right) = -(\alpha_{12}\beta_{11} - \beta_{12}\alpha_{11})^2 < 0$$

Thus, as ω^2 varies from 0 to ∞, $F(\omega^2)$ is first positive, then negative, and then again positive. Hence, it changes sign twice, so that Eq. (7.24) has two positive roots ω_1^2 and ω_2^2, and, as was asserted, all the values for frequency are real.

The quantity ω has four values, both pairs of which are equal in absolute value. If we represent the solution in the form (7.21), it is sufficient to take only positive ω.

The proof that all ω_i^2 are in the most general case positive, provided the quadratic form for $U(q)$ is positive, is analogous, though more involved.

Normal Coordinates. Substitute the roots of the frequency equation into the set of equations (7.22). To each frequency there corresponds a definite ratio of the required quantities A_2/A_1, or in the general case, A_μ/A_1. These ratios are equal to the ratios of the minors of the elements in the first row of determinants (7.23). For the case of two degrees of freedom the ratio $\zeta_i \equiv A_2^{(i)}/A_1^{(i)}$ is apparent directly from Eq. (7.22):

$$\zeta_i = \frac{A_2^{(i)}}{A_1^{(i)}} = -\frac{\beta_{11} - \omega_i^2 \alpha_{11}}{\beta_{12} - \omega_i^2 \alpha_{12}}, \quad i = 1, 2 \tag{7.25}$$

Here the index i corresponds to the number of the solution of the frequency equation (7.23).

Each frequency ω_i defines one partial solution of the set (7.20a) or (7.20b). The general solution of a linear system of equations is

represented as a sum of partial solutions:

$$q_1 = A_1^{(1)}e^{i\omega_1 t} + A_1^{(2)}e^{i\omega_2 t}, \qquad q_2 = \zeta_1 A_1^{(1)}e^{i\omega_1 t} + \zeta_2 A_1^{(2)}e^{i\omega_2 t}$$

$$q_\mu = \sum_h \zeta_{\mu h} A_1^{(h)}e^{i\omega_h t}, \qquad \zeta_{1h} = 1 \tag{7.26}$$

We must, of course, take only the real parts of these expressions. We now introduce the following notation:

$$Q_1 \equiv A_1^{(1)}e^{i\omega_1 t}, \qquad Q_2 \equiv A_1^{(2)}e^{i\omega_2 t}, \ Q_h \equiv A_1^{(h)}e^{i\omega_h t} \tag{7.27}$$

It is immediately apparent from this that Q_1, Q_2, and Q_h satisfy the differential equations

$$\ddot{Q}_1 + \omega_1^2 Q_1 = 0, \qquad \ddot{Q}_2 + \omega_2^2 Q_2 = 0, \qquad \ddot{Q}_h + \omega_h^2 Q_h = 0 \tag{7.28}$$

Each of these equations can be obtained using the Lagrangian

$$L_i = \frac{1}{2}\dot{Q}_i^2 - \frac{1}{2}\omega_i^2 Q_i^2 \tag{7.29}$$

which describes oscillations with one degree of freedom.

Thus, in terms of the variables Q_i, the problem of coupled oscillations with many degrees of freedom is reduced to the problem of independent oscillations of linear harmonic oscillators whose number equals the number of degrees of freedom of the initial oscillating system. Each harmonic oscillator is described in terms of the corresponding coordinate Q_i.

If we substitute the expressions (7.27) into (7.26), we obtain a relationship between the initial generalized coordinates q_μ and the quantities Q_i, which are known as the *normal coordinates*. Thus each generalized coordinate appears as the sum of mutually independent normal coordinates varying according to a harmonic law. Usually the oscillation frequencies ω_i are incommensurable. But then the sum of expressions involving incommensurable frequencies is a nonperiodical function of time.

Whatever the initial conditions, in Eqs. (7.20) it can never be assumed that one of the generalized coordinates remains unaffected by the oscillations. It is sufficient to have a system start oscillating at some instant with respect to even one degree of freedom corresponding to some generalized coordinate, for oscillations to start over all other degrees of freedom. This is due to the mixed components of the Lagrangian involving the products $\dot{q}_\mu \dot{q}_\nu$ or $q_\mu q_\nu$ connecting the degrees of freedom. Conversely, Q_i and $Q_{k \neq i}$ are quite unconnected as long as the solution of the problem takes into account only the quadratic terms in the potential and kinetic energy.

From Eqs. (7.26), Q_1 and Q_2 can be expressed in terms of q_1 and q_2 thus:

$$Q_1 = \frac{\zeta_2 q_1 - q_2}{\zeta_2 - \zeta_1}, \qquad Q_2 = \frac{\zeta_1 q_1 - q_2}{\zeta_1 - \zeta_2} \qquad (7.30)$$

If, for example, the initial values of q_1 and q_2 are so chosen that at that instant of time $Q_1 = 0$ and $\dot{Q}_1 = 0$, the oscillation with frequency ω_1 will not occur at all. For that it is sufficient to take at $t = 0$ the coordinates and velocities in such a proportion that $\zeta_2 q_1 - q_2 = 0$ and $\zeta_2 \dot{q}_1 - \dot{q}_2 = 0$. In other words, only the frequency ω_2 is present, and the oscillations are strictly periodical, which does not occur in the case of arbitrary initial conditions.

It is immediately apparent from the Lagrangian (7.29) that the total energy expression in normal coordinates reduces to the form

$$E = \frac{1}{2} \sum_i (\dot{Q}_i^2 + \omega_i^2 Q_i^2) \qquad (7.31)$$

since $L = T - U$, and $E = T + U$. Thus, the energy of linear harmonic oscillators in terms of the coordinates Q_i replaces the energy of coupled oscillations in terms of the coordinates q_μ.

It should be noted that if the normal coordinates are expressed directly from Eqs. (7.27) or (7.30), the individual energy components are additionally multiplied by certain constant numbers a_i. But if Q_i is replaced by $Q_i (a_i)^{1/2}$, these numbers are eliminated from the energy expression, and it reduces to the form (7.31). An example of such substitution is offered in the exercise at the end of this section.

Thanks to normal coordinates the consideration of oscillation problems is greatly simplified, because the linear harmonic oscillator is in many respects one of the most simple mechanical systems.

The reduction to normal coordinates is essential in studies of the oscillations of polyatomic molecules, in the theory of crystals, and in field theory. In addition, normal coordinates are useful in technical applications of oscillation theory.」

The Case of Equal Frequencies. If the roots of Eq. (7.24) coincide, the general solution must not be written in the form (7.26), but somewhat differently, namely,

$$\left. \begin{array}{l} x_1 = A \cos \omega t + B \sin \omega t \\ y_2 = A' \cos \omega t + B' \sin \omega t \end{array} \right\} \qquad (7.32)$$

Four arbitrary constants appear in this solution, as it should be in a system with two degrees of freedom.

An example of such a system is a pendulum suspended by a string instead of a hinge. In the approximation (7.32), it turns out that the

pendulum describes an ellipse whose axes are tilted with respect to the x and y axes of the coordinate system and whose centre coincides with the origin. If in (7.5) account is taken of the subsequent terms in the potential energy expansion in powers of the deflection angle of the pendulum, we find that the oscillation is accompanied by a rotation of the axes of the ellipse.

EXERCISE

Find the frequencies and normal oscillations of a double pendulum, taking the ratios of load masses $\mu = 3/4$ and rod lengths $\lambda = 5/7$.

Solution. From the equation for the oscillation frequencies of a double pendulum,

$$\omega_1^2 = \frac{7}{2}\frac{g}{l}, \qquad \omega_2^2 = \frac{7}{10}\frac{g}{l}$$

Further, $\zeta_1 = -7/3$ and $\zeta_2 = 7/5$.

Let us now write the expression for kinetic energy. For simplicity, we write $l = g = m = 1$ so that only ratios λ and μ will appear in the equations. This gives $\alpha_{11} = 1 + \mu = 7/4$, $\alpha_{12} = \mu\lambda = 15/28$, $\alpha_{22} = \mu\lambda^2 = 75/196$, $\beta_{11} = 1 + \mu = 7/4$, $\beta_{12} = 0$, and $\beta_{22} = \mu\lambda = 15/28$.

Let us determine the coefficients a_i. To do this we must calculate the kinetic energy:

$$2T = \frac{7}{4}(\dot{Q}_1 + \dot{Q}_2)^2 + \frac{15}{14}(\dot{Q}_1 + \dot{Q}_2)\left(-\frac{7}{3}\dot{Q}_1 + \frac{7}{5}\dot{Q}_2\right)$$
$$+ \frac{75}{196}\left(-\frac{7}{3}\dot{Q}_1 + \frac{7}{5}\dot{Q}^2\right)^2 = \frac{4}{3}\dot{Q}_1^2 + 4\dot{Q}_2^2$$

Consequently, we must put $a_1 = \sqrt{3}/2$ and $a_2 = 1/2$.

Denoting $Q_1\sqrt{3}/2$ and $Q_2/2$ again by he letters Q_1 and Q_2, we have the expression for potential energy

$$2U = \frac{7}{4}\left(\frac{\sqrt{3}}{2}Q_1 + \frac{1}{2}Q_2\right)^2 + \frac{15}{28}\left(-\frac{7}{2\sqrt{3}}Q_1 + \frac{7}{10}Q_2\right)^2$$
$$= \frac{7}{2}Q_1^2 + \frac{7}{10}Q_2^2$$

as it should be according to (7.10). The generalized coordinates are related to the normal coordinates as follows:

$$Q_1 = \frac{5\sqrt{3}}{28}\left(\frac{7}{5}\varphi - \psi\right), \qquad Q_2 = \frac{5\sqrt{3}}{28}\left(\frac{7}{\sqrt{3}}\varphi + \sqrt{3}\psi\right)$$

Thus, if $7\varphi = -3\psi$ and $7\dot{\varphi} = -3\dot{\psi}$ initially, then we have $Q_2 = 0$ for all moments of time, so that both pendulums oscillate with one frequency ω_1, with the constant relationship between the deflection angles $7\varphi = -3\psi$ holding all the time. Both pendulums are deflected to opposite sides of the vertical. The other normal oscillation, with frequency ω_2, occurs for a constant angular relationship $7\varphi = 5\psi$. In this case the pendulums are deflected to the same side.

8

NONINERTIAL FRAMES OF REFERENCE

In Section 3 we stated the principle of equivalence of all inertial frames of reference, or the relativity principle. It is a reflection of the special significance of inertial frames in mechanics.

However, the frames of reference we conventionally treat as inertial are actually only approximations. A frame fixed on the surface of the earth, for example, cannot be considered strictly inertial because of the earth's diurnal rotation. That is the reason why the oscillation plane of a Foucault pendulum rotates with a speed depending on the geographical latitude (see Exercise 1).

The rotation of the oscillation plane of a Foucault pendulum cannot be explained in terms of any interaction with the earth, because gravity cannot make the pendulum rotate precisely from east to west rather than from west to east[2]. However, when only a few oscillations are considered the rotation of the plane is still small and can be neglected. A reference frame fixed on the earth does not have time to display its noninertial qualities. There is always a certain measure of error with which a given real reference frame approximates an inertial system, and it depends on the duration of the motion process being investigated.

Thus, the concept of an inertial reference frame is meaningful as an approximation and is an extremely convenient idealization for mechanics. In such a reference frame the interaction forces are measured in terms of the acceleration of bodies. In the absence of inertial frames we would have to treat the fundamental equation of mechanics (1.1) as merely a definition of force in general and could write it as an identity. Yet it is significant that, thanks to the possibility of observing mechanical systems in inertial frames of reference, Eq. (1.1) makes it possible to determine the measure of physical interactions between bodies. This measure does not depend on the choice of an inertial reference frame, which is precisely what the relativity principle is all about.

The Galilean Transformations. The mathematical expression of the relativity principle consists in that the equations of motion written for one inertial frame retain their form in a transformation to another inertial frame.

---

[2] It is important that the oscillation plane of a Foucault pendulum is strictly vertical, since otherwise the pendulum would have an initial angular momentum around the vertical and describe an ellipse whose axes would necessarily rotate (see end of Section 7).

The equations for transforming from one inertial frame to another can be obtained only on the basis of certain physical assumptions. In Newtonian mechanics it is always assumed that the forces of interaction between bodies, notably gravity, are transmitted instantaneously over any distance. Therefore the displacement of a body immediately imparts a certain momentum to another body, wherever it is located. Thanks to this a clock located in an inertial frame can be immediately synchronized with a clock moving together with another
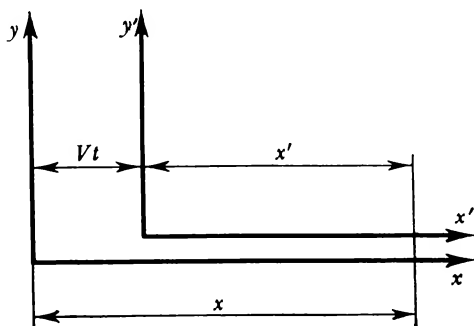
Figure 9

inertial frame. Thus, time in mechanics is accepted as universal; this is not a supplementary hypothesis but a corollary of the assumption of instantaneous action at a distance. In electrodynamics, where the speed of propagation of interactions is finite, time is not universal.

In Newtonian mechanics, however, it is assumed that in going over from one inertial frame of reference to another, moving with a speed V relative to the first, time in both systems is the same, that is, that the readings of initially synchronized clocks always coincide. Later we shall see that the assumption is an approximation valid only when the relative velocities are substantially less than the speed of light.

We shall now obtain the equations for transforming from one inertial frame of reference to another. Let the coordinate axes in both frames be drawn so that the abscissas are directed along the relative velocity V and the ordinates are mutually parallel. We can observe directly from Figure 9 that the abscissa x of a certain point in the frame we agree to call fixed is connected with the abscissa x' in the moving frame by the simple relationship

$$x = x' + Vt \tag{8.1}$$

provided the origins coincided at time $t = 0$.

The other transformation formulas are even simpler:

$$y = y', \qquad z = z' \qquad\qquad\qquad (8.2)$$

The relationship $t = t'$ is ·an expression of the hypothesis of the universality of time; the limits of its application are established by Einstein's relativity principle.

Condition (8.1) is absolutely symmetrical with respect to both inertial reference frames: if the one with the primed quantities is assumed fixed and the other moving, the form of (8.1) does not change, though, of course, V must be replaced by $-V$.

In the present case symmetry is assured by the fact that $t = t'$. If $t \neq t'$, the transformation equations $x = x' + Vt$ and $x' = x - Vt'$ would contradict each other. That is, if time is not assumed the same in all inertial frames, the mathematical expression of the relativity principle should be more complex than Eq. (8.1). But it would appear that (8.1) follows quite obviously from Figure 9. Here we must in large measure forgo the "self-evidence" which in fact stems from our everyday experience with velocities that are small in comparison with the speed of light.

It can easily be shown on the basis of Eq. (8.1) why the Newtonian equations have the same form in all inertial reference frames. The forces of interaction depend on the relative coordinates of the particles; they therefore do not change in the transformation (8.1), since the common term Vt cancels out in the argument of any function involving the differences between the coordinates. The left-hand sides of the Newtonian equations contain accelerations, that is, the second derivatives of the coordinates with respect to time. But since time is involved linearly in Eq. (8.1) and, according to the basic assumption, is the same in both frames of reference, $\ddot{x} = \ddot{x}'$. Hence the equations of mechanics are of identical form in any inertial reference frame. In other words, it is conventionally said that the equations of mechanics are invariant with respect to these transformations, usually known as the *Galilean transformations*.

The invariance of the laws of mechanics in the Galilean transformations is the essence of the relativity principle of Newtonian mechanics. It should be borne in mind here that the relativity principle, which is an expression of the equivalence of all inertial frames of reference, is a reflection of a much more general law of nature than the approximate formula (8.1). In application to electromagnetic phenomena this formula and the equality $t = t'$ are replaced by much more general relationships, from which the Galilean transformations emerge as a limiting case when the velocities of the particles and relative velocities of the reference frames are small in comparison with the speed of light.

Rotating Frames of Reference. The relativity principle does not, of course, imply the equivalence of inertial and noninertial frames of reference.

Notably, in transformations to rotating frames several new characteristic terms appear in the equations of mechanics, which we shall derive later.

Drawing the z axis along the rotation axis, we denote the components of the radius vector $\mathbf{r}'$ in the fixed frame x', y', z', and in the rotating frame, x, y, z. As is known from analytic geometry the components are connected by the equations

$$x = x' \cos \alpha + y' \sin \alpha$$
$$y = -x' \sin \alpha + y' \cos \alpha$$
$$z = z'$$

where α is the angle of rotation.

Differentiate these equations with respect to time and denote the derivative $\dot{\alpha}$ by the letter ω. This is the angular velocity of rotation, or the number of radians per second. We have

$$\dot{x} = \dot{x}' \cos \alpha + \dot{y}' \sin \alpha - \omega x' \sin \alpha + \omega y' \cos \alpha$$
$$= \dot{x}' \cos \alpha + \dot{y}' \sin \alpha + \omega y$$
$$\dot{y} = -\dot{x}' \sin \alpha + \dot{y}' \cos \alpha - \omega x' \cos \alpha - \omega y' \sin \alpha$$
$$= -\dot{x}' \sin \alpha + \dot{y}' \cos \alpha - \omega x$$
$$\dot{z} = \dot{z}'$$

Suppose a line segment of length ω is laid off along the z axis. Considering it as a vector, we have $\omega_z = |\,\boldsymbol{\omega}\,|$, $\omega_x = 0$, $\omega_y = 0$. Having the vector $\boldsymbol{\omega}$, we can replace the product ωy by $-(\boldsymbol{\omega} \times \mathbf{r})_x$, and $-\omega x$ by $-(\boldsymbol{\omega} \times \mathbf{r})_y$. The terms involving $\dot{x}'$, $\dot{y}'$, and $\dot{z}'$ are essentially the projections of the particle's velocity in the inertial frame transformed to the rotating reference frame. Calling them for the time being $\dot{x}'_{\text{rot}}$, $\dot{y}'_{\text{rot}}$, $\dot{z}'_{\text{rot}}$, and expressing them with the help of the obtained equations, we have

$$\dot{x}'_{\text{rot}} = \dot{x} + (\boldsymbol{\omega} \times \mathbf{r})_x$$
$$\dot{y}'_{\text{rot}} = \dot{y} + (\boldsymbol{\omega} \times \mathbf{r})_y$$
$$\dot{z}'_{\text{rot}} = \dot{z} \tag{8.3}$$

These equations can be combined with the help of one vector equation

$$\dot{\mathbf{r}}' = \dot{\mathbf{r}} + \boldsymbol{\omega} \times \mathbf{r} \tag{8.4}$$

Now it is not hard to find the Lagrangian for the variables referring to the rotating reference frame. First, it is obvious that

$$\dot{\mathbf{r}}_{rot}^{'2} = \dot{x}_{rot}^{'2} + \dot{y}_{rot}^{'2} + \dot{z}_{rot}^{'2} = \dot{x}^{'2} + \dot{y}^{'2} + \dot{z}^{'2} = \dot{\mathbf{r}}^{'2}$$

since the absolute value of any vector is the same in any coordinate system. Consequently, in transforming from the variables in the inertial frame to the variables in the rotating frame we obtain the required expression for L:

$$L = \frac{m}{2}\dot{\mathbf{r}}^{'2} - U = \frac{m}{2}\dot{\mathbf{r}}_{rot}^{'2} - U = \frac{m}{2}(\dot{\mathbf{r}} + \boldsymbol{\omega} \times \mathbf{r})^2 - U(r) \qquad (8.5)$$

For the sake of simplicity L is written for one material point.

Now let us write the Lagrange equation for motion relative to the rotating reference frame, that is, taking x, y, and z as the generalized coordinates. Returning for the time being to the components chosen in developing Eq. (8.3), the Lagrangian in terms of the components of the vector takes the form

$$L = \frac{m}{2}[(\dot{x} - \omega y)^2 + (\dot{y} + \omega x)^2 + \dot{z}^2] - U(x, y, z) \qquad (8.6)$$

Whence we obtain

$$\frac{\partial L}{\partial \dot{x}} = m(\dot{x} - \omega y), \qquad \frac{\partial L}{\partial \dot{y}} = m(\dot{y} + \omega x), \qquad \frac{\partial L}{\partial \dot{z}} = m\dot{z}$$

$$\frac{\partial L}{\partial x} = m\omega(\dot{y} + \omega x) - \frac{\partial U}{\partial x}, \qquad \frac{\partial L}{\partial y} = -m\omega(\dot{x} - \omega y) - \frac{\partial U}{\partial y}$$

$$\frac{\partial L}{\partial z} = -\frac{\partial U}{\partial z}$$

The Lagrange equations in terms of the components are

$$m(\ddot{x} - \omega\dot{y}) - m\omega(\dot{y} + \omega x) - m\omega\dot{y} + \frac{\partial U}{\partial x} = 0$$

$$m(\ddot{y} + \omega\dot{x}) + m\omega(\dot{x} - \omega y) + m\omega\dot{x} + \frac{\partial U}{\partial y} = 0$$

$$m\ddot{z} + \frac{\partial U}{\partial z} = 0$$

Let us leave only the second derivatives on the left and rewrite the last three equations as a single vector equation:

$$m\ddot{\mathbf{r}} = m\mathbf{r} \times \dot{\boldsymbol{\omega}} + 2m\dot{\mathbf{r}} \times \boldsymbol{\omega} + m\boldsymbol{\omega} \times (\mathbf{r} \times \boldsymbol{\omega}) - \frac{\partial U}{\partial \mathbf{r}} \qquad (8.7)$$

Expanding the double vector product by means of the equation $\mathbf{A} \times (\mathbf{B} \times \mathbf{C}) = \mathbf{B}(\mathbf{A} \cdot \mathbf{C}) - \mathbf{C}(\mathbf{A} \cdot \mathbf{B})$, and transforming to components, we see that (8.7) is equivalent to the preceding system of

three equations. As always, the advantage of vector notation is that it does not refer the equation to a specific coordinate system and is much more graphic.

Inertial Forces. The first three terms on the right in (8.7) essentially distinguish the equations of motion, written relative to a rotating frame of reference, from the equations written relative to an inertial frame.

The use of a noninertial frame is determined by the nature of the problem. For example, if the motion of terrestrial bodies is being studied, it is natural to choose the earth as the frame of reference, and not the sun. If we consider the reaction of a passenger to a train that suddenly stops, we must take the train as the frame of reference and not the station platform. When the train is braked sharply, the passenger continues to move forwards "inertially" or, more precisely, at the initial time of braking still has the velocity of the train's uniform motion. Thus, relative to the carriage, there is the familiar jerk forward. Obviously, the noninertial frame is the train and not the earth, since no one experiences any jerk on the platform.

The additional terms on the right in Eq. (8.7) have the same origin as the jerk when the train stops—they are produced by the noninertial character (in the given case, rotation) of the frame of reference. Naturally, the acceleration of a point caused by the noninertial character of the reference frame is absolutely real, relative to that frame, in spite of the fact that there are other, inertial, frames relative to which this acceleration does not exist. In equation (8.7) this acceleration is written as if it were due to some additional forces. These forces are usually called *inertial forces*. Unlike interaction forces, inertial forces are proportional only to the mass of the body to which they are applied. This is natural, since the accelerations caused by the noninertial character of a reference frame are, by definition, the same for all bodies placed at the same point of the reference frame and moving in the same way within it. The term "force" is applied to them because the respective expressions are proportional to the product of the mass times the acceleration.

There is only one natural force possessing this property, Newtonian gravitation. The accelerations of all free falling bodies are, as is known, the same. If in Eq. (8.7) we put $U = mgz$, the mass of the moving body cancels out, yielding a universal law of motion that does not depend on the body's mass. The same would refer to the case of $U = 0$. In a noninertial frame, bodies accelerated by inertial forces move neither rectilinearly nor uniformly. Thus there exists a remarkable likeness between a body freely moving in a noninertial frame and a body subject, in addition to the usual forces, to the action of gravity. Einstein's theory of gravitation is based on this experimental fact.

Let us now consider in more detail the inertial forces appearing in (8.7), which are due to the rotation of the reference frame. The first term on the right in (8.7) occurs as a result of the angular velocity being variable. It will not interest us. The second term is called the *Coriolis force*. For a Coriolis force to appear, the velocity of a point relative to a rotating reference frame must have a nonzero projection on a plane perpendicular to the axis of rotation. This velocity projection can, in turn, be resolved into two components: one perpendicular to the radius drawn from the axis of rotation to the moving point, and the other directed along the radius. The most interesting, as to its action, is the component of the Coriolis force due to the radial component of velocity. It is perpendicular both to the radius and to the axis of rotation. If a body moves perpendicular to the radius and to the axis of rotation, then the Coriolis force is radial and is compounded with the centripetal force, which will be considered further on.

We note that the Coriolis force cannot be related, even formally, to the gradient of a potential function.

There are many examples of the action of the Coriolis force in nature. The water of rivers in the Northern Hemisphere that flow in the direction of the meridian, that is, from north to south or from south to north, experiences a deflection towards the right bank (looking in the direction of flow). This is why the right bank of such rivers is steeper than the left. It is easy to form the corresponding component of the Coriolis force. The angular velocity vector of the earth's rotation is directed "upwards" from the north pole. The waters of a river flowing southwards at the middle latitudes of the Northern Hemisphere have a velocity component perpendicular to the earth's axis and directed away from the axis. This means that the Coriolis acceleration of the water, relative to the earth, is in a westerly direction or, relative to a river flowing southwards, to the right. If the river flows in a northerly direction, the deflection will be towards the east, that is, again to the right. In the Southern Hemisphere the deflection is to the left bank.

The Coriolis force substantially affects the motion of air and water masses of the earth, even though it is very small in comparison with the force of gravity. Indeed, the angular velocity of the earth as it makes one rotation in 24 hours is slightly less than 10^{-4} radians per second, while the velocities of particles of water or air may range from 10 to 10^4 cm-s^{-1} (the latter in winds of hurricane force). Hence the Coriolis acceleration may range from 10^{-3} to 1 cm-s^{-2}, or from one-millionth to one-thousandth of the acceleration of gravity.

The Coriolis force also causes the rotation of the plane of oscillation of the Foucault pendulum, which is used to prove the earth's rotation about its axis without resorting to astronomical observations. From the dynamic point of view the choice of the reference

frames to be taken as inertial and rotating is highly relevant. The third vector term in Eq. (8.7) is the usual centrifugal force. Indeed, it is perpendicular to the axis of rotation and in absolute value is equal to

$$| m\omega \times (\mathbf{r} \times \omega) | = m | \omega | | \mathbf{r} \times \omega | = m\omega (\omega r \sin \beta) = m\omega^2 r \sin \beta$$

(8.8)

Here, angle β is the angle between the radius vector $\mathbf{r}$ of the given point and the rotation axis; the origin of the coordinate system lies on this axis. The first equality takes account of the fact that the vectors ω and $\omega \times \mathbf{r}$ are perpendicular to each other, so that the absolute value of the vector product is equal to the product of their absolute values.

But $r \sin \beta$ is equal to the distance from the axis of rotation, so that this force satisfies the conventional definition of a centrifugal force.

EXERCISES

1. Consider the rotation of the oscillation plane of a Foucault pendulum under the action of the earth's rotation about its axis.

Solution. At a given point of the earth we direct the x axis to the north and the y axis to the east. Then, if the vertical angular velocity component $\omega_v = \omega \sin \theta$, where θ is the latitude of the location, and ω is the angular velocity of the earth's rotation, we have the equation of motion

$$\ddot{x} = -\omega_0^2 x - 2\dot{y}\omega_v, \qquad \ddot{y} = -\omega_0^2 y + 2\dot{x}\omega_v, \qquad \omega_0^2 = \frac{g}{l}$$

Multiplying the first equation by y and the second by x and then subtracting, we get

$$\frac{d}{dt}(y\dot{x} - x\dot{y}) = -\frac{d}{dt}(y^2 + x^2)\omega_v$$

Integrating and transforming to polar coordinates ($x = r \cos \varphi$, $y = r \sin \varphi$), we have

$$r^2\dot{\varphi} = r^2\omega_v$$

Whence, after cancelling out r^2, we have

$$\dot{\varphi} = \omega_v$$

which gives the angular velocity of rotation of the oscillation plane.

2. Determine the deflection of a falling body to the east.

Solution. Here the horizontal component $\omega_h = \omega \cos \theta$ of the Coriolis force rather than the vertical is important. For deflection x we obtain

$$\ddot{x} = 2\dot{z}\omega_h = 2gt\omega_h$$

whence

$$x = \frac{1}{3} g t^3 \omega_h$$

Expressed in terms of the height of fall the equation is as follows:

$$x = \frac{1}{3} \frac{(2z)^{3/2}}{g^{1/2}} \omega_h$$

Substitution of real figures reveals that observation of such deflection is difficult.

9

DYNAMICS OF RIGID BODIES

The dynamics of a rigid body is a large independent chapter of mechanics and is very rich in technical applications. Our aim is to give only a brief account of the basic concepts of this branch of mechanics, inasmuch as it contains instructive examples of general laws. In addition, certain mechanical quantities that characterize a rigid body are necessary for an understanding of molecular spectra.

The Kinetic Energy of a Rigid Body. As was shown in Section 2, a rigid body has six degrees of freedom. Three of them relate to the translational motion of the centre of mass in space. The remaining three correspond to rotation of the body relative to its centre of mass.

In Section 4, it was shown that the kinetic energy of a system consists of the kinetic energy of the motion of the whole mass of the body concentrated at the centre of mass and the kinetic energy of the motion of the separate particles relative to the centre of mass. In the case of a rigid body, relative motion reduces to rotation with the angular velocity ω the same for all the particles. Naturally, both the magnitude and the direction of ω may vary with time.

Let us calculate the kinetic energy of rotation of a rigid body. In the general case, the density ρ of the body may not be uniform over the whole volume of the body, and may depend on the coordinates: $\rho = \rho(x, y, z) = \rho(\mathbf{r})$. The mass of an element of volume dV is equal to $dm = \rho(\mathbf{r}) \, dV$. The velocity of rotation $\mathbf{v}$ is, from (8.4), $\omega \times \mathbf{r}$ (we neglect translational motion). Hence, the kinetic energy of unit mass of the body is

$$(\omega \times \mathbf{r})^2 = \omega^2 r^2 \sin^2 \beta = \omega^2 r^2 - \omega^2 r^2 \cos^2 \beta$$
$$= \omega^2 r^2 - (\omega \cdot \mathbf{r})^2$$

The kinetic energy of the whole body is represented by the integral of this quantity over the volume:

$$T = \frac{1}{2} \int \rho \, (\boldsymbol{\omega} \times \mathbf{r})^2 \, dV \qquad (9.1)$$

Expressing the square of the vector product in terms of the components ω_μ and r_ν, we have

$$\omega^2 = \omega_x^2 + \omega_y^2 + \omega_z^2, \quad \boldsymbol{\omega}\mathbf{r} = \omega_x x + \omega_y y + \omega_z z, \quad r^2 = x^2 + y^2 + z^2$$

We now make use of the summation rule:

$$\omega^2 = \omega_\mu \omega_\mu, \quad \boldsymbol{\omega}\mathbf{r} = \omega_\mu x_\mu, \quad r^2 = x_\mu x_\mu$$

In forming the products $r^2\omega^2$ and $(\boldsymbol{\omega}\cdot\mathbf{r})^2$ the summation index should be denoted by different letters so as it would occur nowhere more than twice. Hence

$$(\boldsymbol{\omega} \times \mathbf{r})^2 = \omega_\mu \omega_\mu x_\nu x_\nu - \omega_\mu x_\mu \omega_\nu x_\nu$$

Let us introduce a symbol that will prove highly useful for the subsequent discourse:

$$\delta_{\mu\nu} \equiv \begin{cases} 1 & \text{at} \quad \mu = \nu \\ 0 & \text{at} \quad \mu \neq \nu \end{cases} \qquad (9.2)$$

With the help of this symbol $\omega_\mu \omega_\mu$ can be identically rewritten as

$$\omega_{\mu\mu} \equiv \delta_{\mu\nu} \omega_\mu \omega_\nu$$

and the square of the vector product as

$$(\boldsymbol{\omega} \times \mathbf{r})^2 = \omega_\mu \omega_\nu \, (r^2 \delta_{\mu\nu} - x_\mu x_\nu)$$

The components of the angular velocity of a rigid body are constant over its volume. Consequently they can be taken out from under the integration sign in (9.1), and the kinetic energy is reduced to the form

$$T = \frac{1}{2} \omega_\mu \omega_\nu \int \rho \, (\delta_{\mu\nu} r^2 - x_\mu x_\nu) \, dV \qquad (9.3)$$

All the integrals involved in (9.3) depend only on the shape of the body and the density distribution in it; in a reference frame moving together with the body they do not depend on its motion. Let us now write all six factors of $\omega_\mu \omega_\nu$ in explicit form, remembering that $x_1 = x$, $x_2 = y$, $x_3 = z$, $\delta_{11} = \delta_{22} = \delta_{33} = 1$, $\delta_{12} = \delta_{13} = \delta_{23} = 0$:

$$I_{11} \equiv I_{xx} \equiv \int \rho \, (r^2 - x^2) \, dV = \int \rho \, (y^2 + z^2) \, dV$$

$$I_{22} \equiv I_{yy} \equiv \int \rho \, (r^2 - y^2) \, dV = \int \rho \, (x^2 + z^2) \, dV$$

$$I_{33} \equiv I_{zz} \equiv \int \rho \, (r^2 - z^2) \, dV = \int \rho \, (y^2 + x^2) \, dV$$

and

$$I_{12} \equiv I_{xy} \equiv - \int \rho xy \, dV$$

$$I_{13} \equiv I_{xz} \equiv - \int \rho xz \, dV$$

$$I_{23} \equiv I_{yz} \equiv - \int \rho yz \, dV$$

The quantities with the same subscripts are called the *moments of inertia*, those with different subscripts are the *products of inertia*. Note that $I_{\mu\nu} = I_{\nu\mu}$. Using the notation $I_{\mu\nu}$, we find that the kinetic energy of a rotating solid body can be rewritten in compact form as

$$T = \frac{1}{2} I_{\mu\nu}\omega_\mu\omega_\nu \tag{9.4}$$

Tensors. Before continuing with rigid-body dynamics, let us examine in greater detail the geometric nature of the quantities $I_{\mu\nu}$. Unlike a vector, whose components have single subscripts, $I_{\mu\nu}$ has two subscripts.

It was pointed out before that a vector quantity is characterized by its transformation law in the rotation of a coordinate system. Denoting the cosines of the angles between the old and new axes by the symbols (μ, ν) (see Exercise 3, Section 4), we write the transformation law applied to the vector components:

$$x'_\mu = (\mu, \nu) \, x_\nu$$

Let us now show that the double-subscript quantity $I_{\mu\nu}$ transforms according to the law

$$I'_{\mu\nu} = (\mu, \lambda) \, (\nu, \varkappa) \, I_{\lambda\varkappa}$$

We start with the symbol $\delta_{\mu\nu}$. Consider three unit vectors directed along the axes of the turned coordinate system. For them we have

$$n_1'^{(1)} = 1, \qquad n_2'^{(1)} = 0, \qquad n_3'^{(1)} = 0$$
$$n_1'^{(2)} = 0, \qquad n_2'^{(2)} = 1, \qquad n_3'^{(2)} = 0$$
$$n_1'^{(3)} = 0, \qquad n_2'^{(3)} = 0, \qquad n_3'^{(3)} = 1$$

Form their scalar products

$$\mathbf{n}'^{(1)}\mathbf{n}'^{(1)} = \mathbf{n}'^{(2)}\mathbf{n}'^{(2)} = \mathbf{n}'^{(3)}\mathbf{n}'^{(3)} = 1$$
$$\mathbf{n}'^{(1)}\mathbf{n}'^{(2)} = \mathbf{n}'^{(1)}\mathbf{n}'^{(3)} = \mathbf{n}'^{(2)}\mathbf{n}'^{(3)} = 0$$

For these products we can introduce the concise notation

$$\mathbf{n}'^{(\mu)}\mathbf{n}'^{(\nu)} = \delta_{\mu\nu}$$

A similar equality must hold in any system, notably the un-primed one. Now transform the unit-vector projections from the primed to the unprimed system. From the general formulas of transformations of vectors we obtain

$$n_{\varkappa}^{'(\mu)} = n_{\alpha}^{(\mu)}(\alpha, \varkappa) = (\mu, \varkappa), \qquad n_{\varkappa}^{'(\nu)} = n_{\alpha}^{(\nu)}(\alpha, \varkappa) = (\nu, \varkappa)$$

From these projections we again form scalar products, and we take advantage of the fact that they are the same in any coordinate system. Now we find that the symbol $\delta_{\mu\nu}$ satisfies the equation we had to prove for $I_{\mu\nu}$:

$$\delta_{\mu\nu} = \mathbf{n}'^{(\mu)}\mathbf{n}'^{(\nu)} = n_{\varkappa}^{'(\mu)}n_{\varkappa}^{'(\nu)} = (\mu, \varkappa)(\nu, \varkappa) = (\mu, \lambda)(\lambda, \varkappa)\delta_{\lambda\varkappa}$$

In other words, this means that $\delta_{\mu\nu}$ retains its fundamental property in a transformation to another coordinate system.

The second term under the integral in (9.3) is the product of components, $x_{\lambda}x_{\varkappa}$, each of which is transformed according to a vector law. Therefore their product transforms in the following manner:

$$x_{\mu}'x_{\nu}' = (\mu, \lambda)(\nu, \varkappa)x_{\lambda}x_{\varkappa}$$

Consequently, the integrand in (9.3) transforms with the help of coefficients $(\mu, \lambda)(\nu, \varkappa)$. Since they are constant for each given rotation of the coordinate system, they can be taken outside the integral sign, thus yielding the required transformation law for $I_{\lambda\varkappa}$. All quantities with such a transformation law are called *tensors of rank 2*. According to this terminology a vector should be called a tensor of rank 1, and a scalar has rank zero. Tensors may be of higher rank than 2. The rank is defined by the number of symbols μ, ν, $\ldots$, involved in the transformation law or, what is the same, by the number of indices occurring once in the tensor expression.

If an index is repeated in a tensor expression, it means that summation with respect to it has occurred and the rank of the initial tensor has been reduced by two. Take, for example, a tensor of rank 3, say, $A_{\mu\nu\lambda}$. It transforms according to the law

$$A_{\mu\nu\lambda}' = (\mu, \alpha)(\nu, \beta)(\lambda, \gamma)A_{\alpha\beta\gamma}$$

Let us sum over the indices μ and ν, that is, write them twice on both sides of the equation. Making use of the property of $(\mu, \alpha)(\mu, \beta)$ just proved, we reduce the transformation law written above to the form

$$A_{\mu\mu\lambda}' = (\mu, \alpha)(\mu, \beta)(\lambda, \gamma) = \delta_{\alpha\beta}(\lambda, \gamma)A_{\alpha\beta\gamma} = (\lambda, \gamma)A_{\alpha\alpha\gamma}$$

But this is the transformation law of a vector, that is, a tensor of rank 2, so that the rank of the initial tensor has been reduced by two units, as was asserted.

The kinetic energy expression (9.3)-(9.4) contains all the indices in pairs, that is, it is a tensor of rank zero: it is a scalar, which, of course, is just as it should be.

Angular Momentum of a Rigid Body. Let us now calculate a projection of the angular momentum of a rigid body. From the definition of angular momentum we have

$$M_x = \int \rho \, (\mathbf{r} \times \mathbf{v})_x \, dV = \int \rho \, [\mathbf{r} \times (\boldsymbol{\omega} \times \mathbf{r})]_x \, dV \qquad (9.5)$$

Expanding the double vector product, we reduce M_x to the form

$$M_x = \int \rho \, (\omega_x r^2 - x\boldsymbol{\omega}\mathbf{r}) \, dV$$

$$= \omega_x \int \rho \, (y^2 + z^2) \, dV - \omega_y \int \rho xy \, dV - \omega_z \int \rho yz \, dV \qquad (9.6a)$$

or

$$M_\alpha = I_{\alpha\beta}\omega_\beta \qquad (9.6b)$$

The expression of angular momentum again involves the components of the quantity I, called the *inertia tensor*.

Comparing (9.6a) and (9.4), we see that

$$M_x = \frac{\partial T}{\partial \omega_x} \qquad (9.7)$$

and M_y and M_z appear analogous. In vector form all three equations can be written as

$$\mathbf{M} = \frac{\partial T}{\partial \boldsymbol{\omega}} \qquad (9.8)$$

Equations (9.7) and (9.8) again express the fact that angular momentum is a generalized momentum related to the rotation of a body. In this sense (9.7) is analogous to (5.4). There is, however, a significant difference, for the components ω_α are not the total time derivatives of any quantities. (This will be demonstrated further on in this section.) Therefore ω_x in (9.7) is not altogether similar to $\dot{\varphi}$ in (5.4).

It is apparent from Eq. (9.6a) that in the most general case the angular momentum vector is not parallel to the angular velocity vector. Parallel vectors, as is known, must be linked by a relation ship of the form $M_\alpha = I\omega_\alpha$, where I is a scalar, not a tensor.

Reduction of the Inertia Tensor to the Principal Axes. Nevertheless, to each specific inertia tensor there correspond certain directions in space, such that if the angular velocity vector is directed along them, the angular momentum is in the same direction. As we have just seen, the condition that the angular momentum and angular velocity

vectors are parallel is expressed thus: $M_\alpha = I\omega_\alpha$. Substituting the angular momentum according to the general formula (9.6*b*), we obtain a set of three equations of the form

$$I_{\alpha\beta}\omega_\beta = I\omega_\alpha \qquad (9.9a)$$

or (in terms of the components):

$$\left.\begin{array}{l} I_{11}\omega_1 + I_{12}\omega_2 + I_{13}\omega_3 = I\omega_1 \\ I_{21}\omega_1 + I_{22}\omega_2 + I_{23}\omega_3 = I\omega_2 \\ I_{31}\omega_1 + I_{32}\omega_2 + I_{33}\omega_3 = I\omega_3 \end{array}\right\} \qquad (9.9b)$$

For these linear homogeneous equations to have solutions their determinant must be zero:

$$\begin{vmatrix} I_{11}-I & I_{12} & I_{13} \\ I_{21} & I_{22}-I & I_{23} \\ I_{31} & I_{32} & I_{33}-I \end{vmatrix} = 0 \qquad (9.10)$$

All three roots of this equation are real and positive numbers. This can be shown by reasoning in the same way as in Section 7, when we proved that ω^2 is positive in a system performing small oscillations (for the case of two roots). We shall not give the proof here.

Suppose that in the most general case all three roots of Eq. (9.10) are different. In substituting them into Eq. (9.9*b*) the components ω_1, ω_2, and ω_3 must be taken proportional to the minors of one of the rows, for example the first. Let $\omega_\alpha^{(i)}$ and $\omega_\alpha^{(k)}$ be solutions corresponding to two different values I_i, I_k. (The Latin index numbers the solutions of Eq. (9.10), the Greek index denotes the vector component; here the summation rule refers only to the indices numbering the components.)

Now write Eqs. (9.9*a*) for two different i, k:

$$I_{\alpha\beta}\omega_\beta^{(i)} = I_i\omega_\alpha^{(i)}, \qquad I_{\alpha\beta}\omega_\beta^{(k)} = I_k\omega_\alpha^{(k)}$$

Multiply the left equation by $\omega_\alpha^{(k)}$ and the right one by $\omega_\alpha^{(i)}$, and subtract one from the other to get

$$I_{\alpha\beta}\omega_\alpha^{(k)}\omega_\beta^{(i)} - I_{\alpha\beta}\omega_\alpha^{(i)}\omega_\beta^{(k)} = (I_i - I_k)\,\omega_\alpha^{(i)}\omega_\alpha^{(k)}$$

For the left-hand side of this equation we make use of the basic property of indices over which the summation is performed: any pair of identical indices in any term of the equation can be denoted by another letter without changing the indices in any other terms of the equation. This is due to the fact that the denomination of an index passing from 1 through 3 in summation is quite immaterial. Accordingly, in the second term on the left we redesignate α as β, and β as α. We also take advantage of the fact that $I_{\beta\alpha} = I_{\alpha\beta}$, that is,

the inertia tensor is symmetric with respect to its indices. Hence

$$I_{\alpha\beta}\omega_\alpha^{(i)}\omega_\beta^{(k)} = I_{\beta\alpha}\omega_\beta^{(i)}\omega_\alpha^{(k)} = I_{\alpha\beta}\omega_\alpha^{(k)}\omega_\beta^{(i)}$$

Then the corresponding terms cancel out, and we find that

$$(I_i - I_k)\,\omega_\alpha^{(i)}\omega_\alpha^{(k)} = 0 \tag{9.11}$$

But we agreed in advance that the roots of Eq. (9.10) are different, hence the first term in (9.11) is nonzero. Consequently the scalar product $\omega^{(i)}\omega^{(k)}$ vanishes, and vectors $\omega^{(i)}$ and $\omega^{(h)}$ are mutually perpendicular. Thus, for every point of a rigid body there are three mutually perpendicular lines, such that in a rotation about them the directions of the angular momentum and angular velocity coincide. These three lines are called the *principal axes of inertia* at the given point, and I_1, I_2, I_3 are the *principal moments of inertia.*

Suppose that two of the three principal moments of inertia are equal. According to (9.11) the directions of the corresponding principal axes are perpendicular to the third axis, for which the moment of inertia is not equal to the other two. We draw a plane perpendicular to the third axis through the origin of the coordinate system. Then in a rotation of the body about any line in this plane the angular momentum $\mathbf{M}$ is directed along the rotation axis. Thus, if $I_1 = I_2$, any two mutually perpendicular lines lying in the plane passing through two axes of inertia can be taken as principal axes. In other words, if the axis of rotation lies in the plane of two equal moments of inertia, then the vectors $\mathbf{M}$ and ω are parallel.

If all three principal moments of inertia coincide, $I_1 = I_2 = I_3$, then in a rotation about any axis the angular momentum $\mathbf{M}$ is directed along that axis.

Now write the expression for the kinetic energy of rotation, assuming that the coordinate axes coincide with the principal axes of inertia. From the definition of principal axes of inertia we have

$$M_1 = I_1\omega_1, \qquad M_2 = I_2\omega_2, \qquad M_3 = I_3\omega_3 \tag{9.6c}$$

These equations are obtained if an arbitrary vector of angular momentum M is resolved along the principal axes. The fact that in the most general case each component is multiplied by its number I_i is additional indication that vectors $\mathbf{M}$ and ω are not parallel. We make use of Eqs. (9.7), which yield

$$M_1 = I_1\omega_1 = \frac{\partial T}{\partial \omega_1}, \quad M_2 = I_2\omega_2 = \frac{\partial T}{\partial \omega_2}, \quad M_3 = I_3\omega_3 = \frac{\partial T}{\partial \omega_3}$$

Integrating, we obtain the required expression for the kinetic energy:

$$T = \frac{1}{2}\,(I_1\omega_1^2 + I_2\omega_2^2 + I_3\omega_3^2) \tag{9.12}$$

This equation can be graphically interpreted in the following way. Lay off the angular velocity components ω_1, ω_2, and ω_3 along the axes of an arbitrary coordinate system. Then at constant T Eq. (9.12) represents a triaxial ellipsoid. Knowing the direction of the rotation axis and the kinetic energy of rotation, we can find all three angular velocity components according to the point of intersection of the axis with the ellipsoid. If two moments of inertia are equal, the triaxial ellipsoid becomes an ellipsoid of rotation; if three moments of inertia are equal, it becomes a sphere. It is clear from this why the choice of principal axes of inertia in a plane of equal moments of inertia is arbitrary.

Euler's Equations of Motion for a Rotating Rigid Body. We shall now develop equations that show how the angular momentum **M** varies with time. For a separate mass point we have

$$\frac{d\mathbf{M}}{dt} = \frac{d}{dt}(\mathbf{r} \times \mathbf{p}) = \dot{\mathbf{r}} \times \mathbf{p} + \mathbf{r} \times \dot{\mathbf{p}} = \mathbf{r} \times \mathbf{F}$$

where the first term vanishes, since $\dot{\mathbf{r}}$ and $\mathbf{p}$ are parallel. Integrating this equation over the volume of the rigid body and taking advantage of the additive property of angular momentum, we have

$$\dot{\mathbf{M}} = \int (\mathbf{r} \times \mathbf{F})\, dV = \mathbf{K} \qquad (9.13)$$

The vector in the right-hand side of (9.13), which we denote by **K**, is called the *resultant torque on the body*. If **F** is the gravitational force (which occurs in the majority of cases), then torque **K** can also be written as

$$\mathbf{K} = -\int \rho g\, (\mathbf{r} \times \mathbf{z}_0)\, dV$$

where $\mathbf{z}_0$ is the unit vector in a vertical direction. But since $\mathbf{z}_0$ is a constant, it should be taken outside the integration sign, and for the torque due to gravitational forces we obtain the expression

$$\mathbf{K} = \mathbf{z}_0 \times \int \rho g \mathbf{r}\, dV$$

Suppose a body is supported at its centre of mass or, what is the same thing, at its centre of gravity. Then the resultant gravitational force is balanced by the reaction of the support, and the resultant torque vanishes, since from the definition of the centre of mass the integral for all three projections is zero. If $\mathbf{K} = 0$, the rigid body rotates as a free body. This case of motion occurs in a demonstration gyroscope.

For the angular momentum of a rigid body to be conserved all that is required is for the resultant torque to be zero. The angular

momentum vector of an arbitrary mechanical system is conserved if there are no external forces at all (with the exception of the case of the whole system moving in the field of a fixed attractive centre).

Equation (9.13) is inconvenient for the case of a fixed frame of reference. In such a system the inertia tensor linking vectors **M** and **ω**, according to Eq. (9.6*b*), itself becomes a variable, because the integrals involved in Eq. (9.3), taken with respect to a fixed frame, are in the most general case variable over the volume of the rotating rigid body. It is preferable to refer the equation to a reference frame fixed relative to the body, taking into account the frame's accelerated motion. The change of vector **M** relative to the moving axes consists of two components: one due to the change of the vector itself, the other to the motion of the axes on which it is projected. For vector **M** the latter change is equal to **ω × M**, just as for the radius vector **r** in Section 8 it was equal to **ω × r**. (In a rotation of a coordinate system any vector varies as a radius vector.) Then Eq. (9.13) written with respect to moving axes looks like this:

$$\frac{d\mathbf{M}}{dt} + \boldsymbol{\omega} \times \mathbf{M} = \mathbf{K} \tag{9.14}$$

where the vector of torque should also be projected on the moving axes.

What, consequently, is required is a set of kinematic equations defining the direction of the moving axes relative to the fixed ones. Two systems of axes can be related by means of nine cosines or, in our notation, symbols $(\mu, \varkappa)$. As was shown, this symbol is the projection of a unit vector $\mathbf{n}^{(\varkappa)}$ on the moving axis numbered μ. The unit vector $\mathbf{n}^{(\varkappa)}$ does not change relative to the fixed axes: $\dot{\mathbf{n}}^{(\varkappa)} = 0$. Writing the derivative $\dot{\mathbf{n}}^{(\varkappa)}$ relative to the moving axes, we obtain

$$\frac{d\mathbf{n}^{(\varkappa)}}{dt} + \boldsymbol{\omega} \times \mathbf{n}^{(\varkappa)} = 0 \tag{9.15}$$

Thus, for the nine cosines we have three vector equations, that is, nine equations. Let us show that these equations do not violate the basic property of unit vectors: $\mathbf{n}^{(\varkappa)} \mathbf{n}^{(\lambda)} = \delta_{\varkappa\lambda}$. For this write Eqs. (9.15) for two different unit vectors $\mathbf{n}^{(\varkappa)}$ and $\mathbf{n}^{(\lambda)}$:

$$\frac{d\mathbf{n}^{(\varkappa)}}{dt} + \boldsymbol{\omega} \times \mathbf{n}^{(\varkappa)} = 0$$

$$\frac{d\mathbf{n}^{(\lambda)}}{dt} + \boldsymbol{\omega} \times \mathbf{n}^{(\lambda)} = 0$$

Multiply the first by $\mathbf{n}^{(\lambda)}$ scalarly, the second by $\mathbf{n}^{(\varkappa)}$, and add. After a cyclic permutation in the mixed products, we obtain

$$\mathbf{n}^{(\lambda)}\,\frac{d\mathbf{n}^{(\varkappa)}}{dt} + \mathbf{n}^{(\varkappa)}\,\frac{d\mathbf{n}^{(\lambda)}}{dt} = \frac{d}{dt}\,\mathbf{n}^{(\lambda)}\mathbf{n}^{(\varkappa)}$$

$$= -\,\boldsymbol{\omega}\,(\mathbf{n}^{(\varkappa)}\times|\mathbf{n}^{(\lambda)}) - \boldsymbol{\omega}\,(\mathbf{n}^{(\lambda)}\times|\mathbf{n}^{(\varkappa)}) = 0$$

since $\mathbf{n}^{(\varkappa)} \times \mathbf{n}^{(\lambda)} = -\mathbf{n}^{(\lambda)} \times \mathbf{n}^{(\varkappa)}$. Hence, if the condition $\mathbf{n}^{(\varkappa)}\mathbf{n}^{(\lambda)} = \delta_{\varkappa\lambda}$ was satisfied at the initial time, it holds subsequently, as was asserted.

Equations (9.6c), (9.14) and (9.15) fully define the position of a rigid body in space.

It is convenient to eliminate the angular momentum components at once with the help of Eqs. (9.6c), thereby referring the motion to the principal axes of inertia, which are fixed relative to the body. Then instead of Eq. (9.14) we obtain the following set:

$$\left.\begin{aligned} I_1\dot{\omega}_1 + (I_3 - I_2)\,\omega_2\omega_3 &= K_1 \\ I_2\dot{\omega}_2 + (I_1 - I_3)\,\omega_3\omega_1 &= |K_2 \\ I_3\dot{\omega}_3 + (I_2 - I_1)\,\omega_1\omega_2 &= K_3 \end{aligned}\right\} \tag{9.16}$$

These equations, known as *Euler's equations*, can be reduced to quadratures for arbitrary values of the integrals of the motion in the following main cases:

(i) The point of support of the body lies at the centre of mass, so that $K_1 = K_2 = K_3 = 0$; the relationship between the principal moments of inertia is arbitrary.

(ii) $I_1 = I_2 \neq I_3$, and the point of support lies on the symmetry axis relative to which two moments of inertia are equal; the point of support does not coincide with the centre of mass. This is known as a symmetric top.

(iii) $I_1 = I_2 = 2I_3$; the centre of mass lies in a plane through the point of support perpendicular to the axis of symmetry. This is *Kovalevskaya's top*.

It has also been shown that the set of equations (9.16) can be integrated in quadratures with arbitrary integrals of the motion only if conditions (i), (ii), or (iii) hold, and in a few similar cases.

The Free Top. Case (i), when $\mathbf{K} = 0$, is integrated analytically in very complex form. We shall therefore consider only some general properties of the rotation of a free top, which can be obtained by investigating the integrals of the motion.

In a fixed frame of reference the angular momentum vector $\mathbf{M}$ is conserved. In a noninertial frame fixed with respect to the top, $\mathbf{M}$ is, of course, not conserved. But since the rotation does not affect

the absolute value of the vector, the square of the angular momentum, M^2, must be conserved in a rotating system. This is easily proved directly by multiplying the first equation in the set (9.16) by $I_1\omega_1$, the second by $I_2\omega_2$, and the third by $I_3\omega_3$. Then, adding all three equations, at $K = 0$ we obtain

$$\frac{d}{dt}(I_1^2\omega_1^2 + I_2^2\omega_2^2 + I_3^2\omega_3^2) = \frac{d}{dt}(M_1^2 + M_2^2 + M_3^2) = \frac{dM^2}{dt} = 0$$

It follows from this set of equations that the kinetic energy T is also conserved. The equations should be multiplied by ω_1, ω_2, ω_3 and added, yielding $T = $ constant.

Let us now express the projections of the angular velocity vectors in terms of the angular momentum projections with the aid of (9.6c) and substitute them into the expression for kinetic energy. Then (together with the condition that the square of the angular momentum is conserved), we obtain two equations:

$$\frac{M_1^2}{2I_1} + \frac{M_2^2}{2I_2} + \frac{M_3^2}{2I_3} = T \tag{9.17}$$

$$M_1^2 + M_2^2 + M_3^2 = M^2 \tag{9.18}$$

If segments M_1, M_2, and M_3 are laid off along the axes of a coordinate system, then (9.18) is the equation of a sphere and (9.17) is the equation of a triaxial ellipsoid. Vector **M** must satisfy both equations, that is, it should lie on the intersection of both planes.

The intersection lines are different depending on where they lie. Their geometry is shown approximately in Figure 10. Near the major axis the ellipsoid intersects with the sphere, and the intersection line is consequently a closed curve. Near the minor axis the ellipsoid is flattened most, so that the sphere emerges from it, and the intersection lines are again closed. Near the median axis there is a section of the ellipsoid with greater curvature than the sphere and one with smaller curvature. As a result the intersection line through the median axis is a cross, while close to it the lines are like hyperbolas, with the crossed lines as their asymptotes.

It follows from this construction that if the angular momentum vector lies somewhere near the major inertia axis, it describes a closed curve around it. In a fixed frame of reference, where the angular momentum vector does not change, the major axis of the ellipsoid rotates around the angular momentum vector along a closed curve. Hence rotation about the major axis of inertia of a body is stable. The same is true of rotation about the minor axis of inertia. As for the median axis, the curves are open, so that the angular momentum vector does not stay close to the axis, and rotation about it is unstable.

For the subsequent investigation it is convenient to lay off the components of the angular velocity vector ω_1, ω_2, and ω_3 along the coordinate axes rather than the angular momentum components.
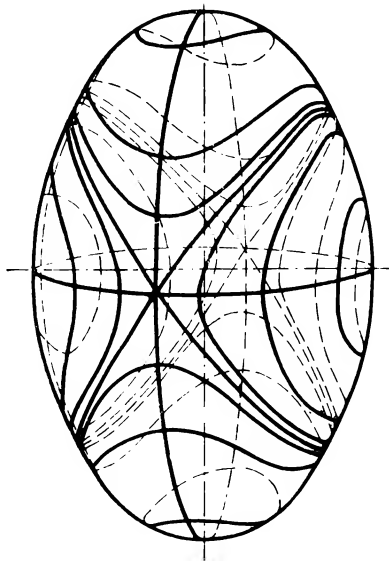

Figure 10

We write the energy and angular momentum equations as follows:

$$\frac{I_1\omega_1^2}{2} + \frac{I_2\omega_2^2}{2} + \frac{I_3\omega_3^2}{2} = T$$

$$I_1^2\omega_1^2 + I_2^2\omega_2^2 + I_3^2\omega_3^2 = M^2$$

Let us find the direction cosines of the normal to the ellipsoid, the ellipsoid which expresses the conservation of energy. We obtain

$$\cos\alpha_1 = \frac{I_1\omega_1}{\sqrt{I_1^2\omega_1^2 + I_2^2\omega_2^2 + I_3^2\omega_3^2}} = \frac{I_1\omega_1}{M} = \frac{M_1}{M}$$

and analogously for the other components. But the ratio M_i/M is the cosine of the angle between the angular momentum vector in the fixed frame of reference, where it is conserved, and the moving axes connected with the body. Consequently, an osculating plane to the ellipsoid remains constantly oriented in space perpendicular to the angular momentum vector.

Now take vector ω drawn towards the point of contact of the plane and the ellipsoid. Its projection on the direction of the normal to the

ellipsoid is

$$d = \omega_1 \cos \alpha_1 + \omega_2 \cos \alpha_2 + \omega_3 \cos \alpha_3 = \sum_i \omega_i \frac{I_i \omega_i}{M}$$

$$= \frac{2T}{M} = \text{constant}$$

This is the projection of a vector, drawn to the point of contact, on the direction of the normal to the surface of the ellipsoid at that point. Hence, it is also the projection of vector ω on the direction of the normal to the osculating plane or, what is the same thing, the length of a perpendicular from the origin of the coordinate system to the osculating plane.

Thus, a plane perpendicular to the total angular momentum and tangent to the ellipsoid whose equation expresses the conservation of energy, remains at a constant distance d from the origin of the coordinate system along the axes of which are laid off the segments ω_1, ω_2, and ω_3, that is, the projections of the angular velocity of the body on the moving axes. The ellipsoid, so to say, rolls along a spatially fixed plane, and the radius vector drawn to the point of contact of the ellipsoid and the plane states the instantaneous value of the angular velocity in both magnitude and direction.

Free Symmetric Top. As pointed out before, the problem on the rotation of a symmetric top whose point of support lies on the axis of symmetry has an exact solution. Mathematically it is rather complex. On the other hand, the solution of the problem of a free symmetric top is quite simple. We shall obtain it in the way just elaborated for the case of an arbitrary free top.

Suppose that $I_2 = I_3$, that is, the ellipsoid whose equation expresses the conservation of energy is an ellipsoid of rotation. It is shown in Figure 11, where the osculating plane has, quite arbitrarily, been chosen horizontal. For the sake of making the drawing clearer, the angular velocity vector has been drawn away from, rather than towards, the point of contact of the ellipsoid and the plane. This simply means that an identical osculating plane is presumed drawn above the ellipsoid. Furthermore, the vectors have not been drawn to scale and are somewhat exaggerated, so that vector ω should be pictured as drawn only up to the surface of the ellipsoid, and its projections reduced accordingly.

It is apparent from Figure 11 that angle θ between the axis of symmetry and the direction of the normal to the plane, that is the direction of the total angular momentum, remains unchanged, because the point of contact describes a circle in a plane perpendicular to the axis of symmetry. Angle θ is determined from the equation

$$\cos \theta = I_1 \omega_1 / M$$

The angular velocity component along the axis of symmetry remains constant. This means that the axis of symmetry revolves uniformly about the direction of the total angular momentum, which is constant in space. Such motion is known as *precession*.

Since precession takes place about the vector of total angular momentum, its angular velocity ω_p must be directed along **M**. The component along the axis of symmetry has no relationship to precession. As can be seen from Figure 11, $\omega_2 = \omega_p \sin \theta$. But the
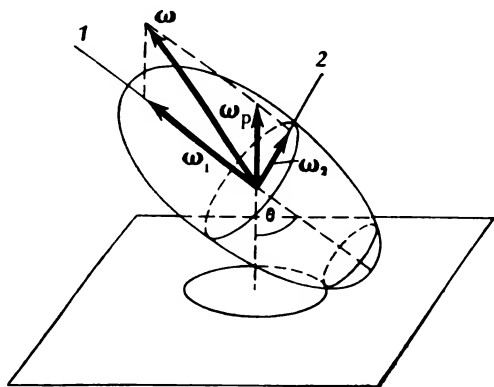
Figure 11

projection of the total angular momentum on axis *2*, which is assumed constant in the plane of the drawing (and not fixed to the body of the top), is equal to $M_2 = I_2 \omega_2 = M \sin \theta$, whence

$$\omega_p = M/I_2 \tag{9.19}$$

If the point of support of the top is not at the centre of mass, the total angular momentum is not conserved, only its vertical component is. At sufficiently high angular velocities the top performs oscillations in the vertical plane, which superimpose on the precession.

The Euler Angles. We shall now show how to describe the rotation of a rigid body in space in terms of the parameters defining its position. For this it is useful to draw two coordinate systems: one fixed, $Oxyz$, and the other fixed with respect to the body, $Ox'y'z'$, usually in such a way that the latter's axes coincide with the principal axes of inertia at the given point. Then the position of the moving coordinate system relative to the fixed one is uniquely given by the three Euler angles (Figure 12):

ϑ is the angle between axes Oz and Oz';

φ is the angle between the intersection line OK of planes xOy and $x'Oy'$ and the x' axis;

ψ is the angle between OK and the x axis.

As we know, the angular velocity of rotation is laid off as a segment directed along the rotation axis, that is, perpendicular to the plane of rotation. Line OK is perpendicular (by the construction) to the z and z' axes, hence it is perpendicular to the plane through them. The angle of rotation ϑ is laid off in this plane, so that the angular velocity $\dot{\vartheta}$ is directed along OK. Similarly, we see that the angular velocity $\dot{\varphi}$ is laid off along the z' axis, and $\dot{\psi}$ along the z axis.
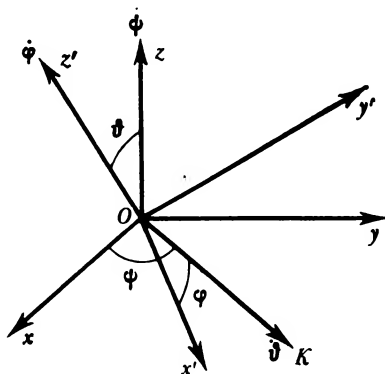
Figure 12

Now express the angular velocity projections ω_1, ω_2, and ω_3 on the principal axes of inertia Ox', Oy' and Oz' in terms of the generalized velocities $\dot{\psi}$, $\dot{\varphi}$, and $\dot{\vartheta}$. The projection ω_3 is that of the angular velocity on the z' axis (the third axis). As pointed out, $\dot{\varphi}$ is projected fully on this axis, while the projection of $\dot{\psi}$ is equal to $\dot{\psi} \cos \vartheta$, since ϑ is the angle between the z and z' axes. Hence

$$\omega_3 = \dot{\varphi} + \dot{\psi} \cos \vartheta \qquad (9.20)$$

In order to find the projections of the angular velocity on the other two axes we mentally draw a line OL in the plane $x'Oy'$ perpendicular to OK (OL is not shown in Figure 12). Then

$$\angle LOx' = \frac{\pi}{2} - \varphi \quad \text{and} \quad \angle zOL = \frac{\pi}{2} + \vartheta$$

since, like all lines perpendicular to OK, OL lies in the z,z'-plane.

The projection of $\dot{\psi}$ on OL is equal to $-\dot{\psi} \sin \vartheta$, and the projection on Ox' is equal to $-\dot{\psi} \sin \vartheta \cos (\pi/2 - \varphi)$, or $-\dot{\psi} \sin \vartheta \sin \varphi$.

The projection of $\dot{\psi}$ on Oy' is $\dot{\psi} \sin \vartheta \cos \varphi$. The projections of $\dot{\vartheta}$ on Ox' and Oy' are apparent from the diagram: they are $\dot{\vartheta} \cos \varphi$ and $\dot{\vartheta} \sin \varphi$. The result is therefore

$$\omega_1 = \dot{\vartheta} \cos \varphi - \dot{\psi} \sin \vartheta \sin \varphi \tag{9.21}$$

$$\omega_2 = \dot{\vartheta} \sin \varphi + \dot{\psi} \sin \vartheta \cos \varphi \tag{9.22}$$

From Eqs. (9.20)-(9.22) we see that ω_1, ω_2, and ω_3 are not total time derivatives of any quantities and, in that sense, do not exactly agree with the usual notion of generalized velocities (as do $\dot{\varphi}$, $\dot{\psi}$, $\dot{\vartheta}$). This was discussed in connection with Eq. (9.7).

If we substitute into (9.12) the expressions for ω_1, ω_2, ω_3 in terms of the Euler angles, we obtain the kinetic energy of a rigid body as a function of the generalized coordinates φ, ψ, ϑ and velocities $\dot{\varphi}$, $\dot{\psi}$, $\dot{\vartheta}$.

A Symmetric Top in a Gravitational Field. We shall find the Lagrangian for a symmetric top whose point of support lies on the axis of symmetry at a distance l below the centre of mass. If the top is inclined at an angle ϑ to the vertical, the height of the centre of mass above the point of support z, is $l \cos \vartheta$. Hence, the potential energy of the top is

$$U = mgz = mgl \cos \vartheta \tag{9.23}$$

The kinetic energy of the top, expressed in terms of the Euler angles, is

$$\begin{aligned} T &= \frac{1}{2} I_1 (\omega_1^2 + \omega_2^2) + \frac{1}{2} I_3 \omega_3^2 \\ &= \frac{1}{2} I_1 (\dot{\vartheta}^2 + \dot{\psi}^2 \sin^2 \vartheta) + \frac{1}{2} I_3 (\dot{\varphi} + \dot{\psi} \cos \vartheta)^2 \end{aligned} \tag{9.24}$$

The difference gives the Lagrangian for a symmetric top. But since it does not contain time explicitly, the total energy $E = T + U$ is an integral of motion:

$$E = T + U = \text{constant}$$

We can find two more integrals of the motion, noting that the angles φ and ψ do not appear explicitly in the Lagrangian, that is, φ and ψ are cyclic coordinates (φ is eliminated only in the case of

a symmetric top, and ψ is not involved in T at all). We obtain

$$p_\varphi = \frac{\partial L}{\partial \dot\varphi} = I_3 (\dot\varphi + \dot\psi \cos\vartheta) = \text{constant} \tag{9.25}$$

$$p_\psi = \frac{\partial L}{\partial \dot\psi} = I_1 \sin^2\vartheta\,\dot\psi + I_3 \cos\vartheta\,(\dot\varphi + \dot\psi \cos\vartheta) = \text{constant} \tag{9.26}$$

If we eliminate $\dot\varphi$ and $\dot\psi$ from the last two equations and substitute them into the energy integral, the latter will contain only the variable ϑ, which allows us to reduce the problem to quadrature.

Substituting (9.25) into (9.26), we obtain

$$p_\psi = I_1 \sin^2\vartheta\,\dot\psi + p_\varphi \cos\vartheta$$

whence

$$\dot\psi = \frac{1}{I_1 \sin^2\vartheta}(p_\psi - p_\varphi \cos\vartheta)$$

The energy integral, after substituting $\dot\varphi$ and $\dot\psi$, is

$$E = \frac{1}{2} I_1\dot\vartheta^2 + \frac{(p_\psi - p_\varphi \cos\vartheta)^2}{2I_1 \sin^2\vartheta} + \frac{p_\varphi^2}{2I_3} + mgl\cos\vartheta \tag{9.27}$$

Thus, the problem is reduced to motion as it were with one degree of freedom ϑ. The corresponding "kinetic energy" is $I_1\dot\vartheta^2/2$, and the "potential energy" is represented by those energy terms which depend on ϑ. This potential energy becomes infinite for $\vartheta = 0$ and for $\vartheta = \pi$. Hence, for $0 < \vartheta < \pi$ it has at least one minimum. If this minimum corresponds to $\vartheta < \pi/2$, the rotation of a top whose centre of mass is above the point of support is stable. Small oscillations are possible around the point of stable equilibrium, which is in this case the dynamic point of equilibrium. These oscillations are superimposed on the precessional motion of the top, which we have already noted. They are called *nutations*.

EXERCISES

1. Knowing the integrals of the motion p_φ and p_ψ, determine the angle at which the amplitude of nutations may vanish, that is, the top rotates in a gravitational field like a free top (pseudoregular precession).

Solution. From the energy integral (9.27) we find the essentially positive quantity $I_1\dot\vartheta^2/2$. In nutations angle ϑ varies but slightly near the position of minimum "potential" energy. If the total energy E corresponds to the

minimum "potential" energy position, both points of intersection of the line $E =$ constant with the "potential" energy curve merge. Differentiating equation

$$E - \frac{(p_\psi - p_\varphi \cos \vartheta)^2}{2I_1 \sin^2 \vartheta} - \frac{p_\varphi^2}{2I_3} - mgl \cos \vartheta = 0$$

with respect to ϑ, equating the derivative to zero, and solving simultaneously with the energy equation, we obtain the equation for the angle of pseudo-regular precession and E:

$$3mgl \cos^2 \vartheta - \left(2E - \frac{p_\varphi^2}{I_3} + \frac{p_\varphi^2}{I_1} \right) \cos \vartheta - mgl + \frac{p_\varphi p_\psi}{I_1} = 0$$

2. The axis of a symmetric top (gyroscope) can rotate only in the horizontal plane. Determine its motion, taking into account the effect of the earth's diurnal rotation.

Solution. Resolve the vector of the angular velocity of the earth's rotation at the given point of the globe into two components: vertical $\omega_v = \omega \sin \theta$ (where θ is the latitude of the location) and horizontal, tangent to the meridian, $\omega_h = \omega \cos \theta$; the latter, obviously, is directed to the north. Let the gyroscope's axis make an angle φ to the meridian in the horizontal plane, and denote the angular rotation of the gyroscope about its axis as ω_0. Since two moments of inertia of the gyroscope are equal, one of the principal axes can be assumed constantly directed vertically, and the other horizontally. Though the gyroscope moves relative to them they do not lose the property of principal axes, because they remain constantly in the plane of equal moments of inertia.

The projections of the angular velocities of rotation of the gimbals of the gyroscope (which keep its axis in the horizontal plane) on the principal axes of inertia thus chosen are:

$$\omega_1 = \omega_v + \dot{\varphi}, \qquad \omega_2 = \omega_h \sin \varphi, \qquad \omega_3 = \omega_h \cos \varphi$$

From this we obtain the angular momentum components along the principal axes of inertia (the angular velocity ω_0 of the gyroscope itself was not taken into account in ω_3, because the reference frame is fixed relative to the gimbals!):

$$M_1 = I_1 (\omega_v + \dot{\varphi}), \qquad M_2 = I_1 \omega_h \sin \varphi$$
$$M_3 = I_3 (\omega_0 + \omega_h \cos \varphi)$$

Now we make use of Eqs. (9.14) to vary the components of **M** with time. The torque acts only relative to the second axis, due to the reaction of the support that keeps the gyroscope's axis in the horizontal plane. There is no need to write this equation. The equation for the first component of angular

momentum yields

$$\frac{dM_1}{dt} + \omega_2 M_3 - \omega_3 M_2$$

$$= I_1 \frac{d^2\varphi}{dt^2} + I_3 \omega_h \sin\varphi \, (\omega_0 + \omega_h \cos\varphi) - I_1 \omega_h^2 \cos\varphi \sin\varphi = 0$$

Neglecting the small quantity of the order of the square of the earth's angular velocity, we arrive at an equation that coincides with the equation of pendulum oscillations, which follows from (3.22):

$$\frac{d^2\varphi}{dt^2} + \frac{I_3}{I_1} \, \omega_0 \omega_h \sin\varphi = 0$$

At small angles of deflection from the meridian, $\sin\varphi$ can be replaced by φ, and the oscillation period is equal to

$$2\pi \left(\frac{I_1}{I_3 \omega_0 \omega \cos\theta} \right)^{1/2}$$

This is the principle of action of the gyroscopic compass which, unlike the magnetic compass, points straight north.

3. Using Euler's equations (9.16) and the kinematic equations (9.15), obtain the first integral of the motion for a Kovalevskaya top.

Solution. Let the point of support of the top lie in the plane of equal moments of inertia at a distance l from the centre of mass. Denoting the projections of the vertical direction on the moving axes n_1, n_2, and n_3, we find that the torque due to the force of gravity relative to the centre of mass has components 0, $-n_3 \, (mg)^{1/2}$, and $n_2 \, (mg)^{1/2}$. We write the kinematic equations for n_1, n_2, and n_3:

$$\frac{dn_1}{dt} + \omega_2 n_3 - \omega_3 n_2 = 0$$

$$\frac{dn_2}{dt} + \omega_3 n_1 - \omega_1 n_3 = 0$$

$$\frac{dn_3}{dt} + \omega_1 n_2 - \omega_2 n_1 = 0$$

We write Euler's equations as follows:

$$2\frac{d\omega_1}{dt} - \omega_2 \omega_3 = 0$$

$$2\frac{d\omega_2}{dt} + \omega_1 \omega_3 = -\mu^2 n_3$$

$$\frac{d\omega_3}{dt} = \mu^2 n_2, \qquad \mu^2 \equiv \frac{mgl}{I_1}$$

Multiply the second equation by i, add to the first, and multiply the result by $\omega_1 + i\omega_2$ to get

$$\frac{d}{dt}(\omega_1 + i\omega_2)^2 = -i\omega_3 (\omega_1 + i\omega_2)^2 - i\mu^2 n_3 (\omega_1 + i\omega_2)$$

Multiply the second kinematic equation by i and add to the first to get

$$\frac{d}{dt}(n_1 + in_2) = -i\omega_3(n_1 + in_2) - in_3(\omega_1 + i\omega_2)$$

Finally, multiply both sides of the obtained relationship and subtract from the equation containing the derivative $(d/dt)(\omega_1 + i\omega_2)^2$. As a result we obtain

$$\frac{d}{dt}[(\omega_1 + i\omega_2)^2 - \mu^2(n_1 + in_2)] = -i\omega_3[(\omega_1 + i\omega_2)^2 - \mu^2(n_1 + in_2)]$$

or, denoting $\theta \equiv (\omega_1 + i\omega_2)^2 - \mu^2(n_1 + in_2)$,

$$\frac{d\theta}{dt} = -i\omega_3\theta$$

The equation for the complex-conjugate quantity θ^* is derived in quite the same way: $d\theta^*/dt = i\omega_3\theta^*$. But then, after multiplying the equation for θ by θ^*, and the one for θ^* by θ, and adding, we obtain the integral of motion:

$$\frac{d}{dt}\theta\theta^* = 0, \qquad |(\omega_1 + i\omega_2)^2 - \mu^2(n_1 + in_2)|^2 = \text{constant}$$

10

HAMILTON'S EQUATIONS AND THE HAMILTON-JACOBI EQUATION

Up till now we used the Lagrange set of equations as the equations of motion. Making use of the concept of generalized momentum p_α, we can write these equations as follows:

$$\frac{dp_\alpha}{dt} - \frac{\partial L}{\partial q_\alpha} = 0 \tag{10.1}$$

$$p_\alpha = \frac{\partial L}{\partial \dot{q}_\alpha} \tag{10.2}$$

The generalized coordinates and momenta are involved non-symmetrically. In some cases it is convenient to have a symmetric set of equations. It is neither better nor worse than the Lagrange equations for practical solutions of mechanical problems, but is of much greater importance in general research.

Hamilton's Equations. The Lagrangian depends on the generalized velocities quadratically, so that Eqs. (10.2) are linear with respect

to all the $\dot{q}_\alpha$'s. Such equations can always be solved and $\dot{q}_\alpha$ expressed in terms of the generalized momenta.

Let us now form the expression for energy (see (4.1)):

$$E = \dot{q}_\alpha \frac{\partial L}{\partial \dot{q}_\alpha} - L(\ldots q_\alpha \ldots, \ldots \dot{q}_\alpha \ldots)$$

and substitute into it all the $\dot{q}_\alpha$'s expressed in terms of all p_β's.

The energy expressed in terms of the generalized coordinates and corresponding momenta is called the *Hamilton function*, or simply the *Hamiltonian*, of the system:

$$E(\ldots q_\alpha \ldots, \ldots \dot{q}_\alpha(p_\beta) \ldots) \equiv \mathscr{H}(p, q) = \dot{q}_\alpha p_\alpha - L \tag{10.3}$$

If, for example, we replace $\dot{\vartheta}$ in expression (9.27) by p_ϑ/I_1, we obtain the Hamiltonian of a symmetric top in a gravitational field.

To develop the required set of equations we write the expression for *Hamilton's principle*, or the *principle of least action*, expressing L in the integrand in terms of H:

$$\delta S = \delta \int_{t_0}^{t_1} [p_\alpha \dot{q}_\alpha - \mathscr{H}(p, q)] \, dt = 0 \tag{10.4}$$

Here all the generalized velocities have been replaced in terms of the generalized momenta. Obviously, for an integral to be extremal along an actual path it cannot depend on the variables in terms of which it is expressed.

Let us calculate the variation δS:

$$\delta S = \int_{t_0}^{t_1} \left(\delta p_\alpha \dot{q}_\alpha + p_\alpha \delta \dot{q}_\alpha - \frac{\partial \mathscr{H}}{\partial p_\alpha} \delta p_\alpha - \frac{\partial \mathscr{H}}{\partial q_\alpha} \delta q_\alpha \right) dt = 0$$

The second term in the parentheses can be integrated by parts as was done in (2.17). Then

$$\delta S = p_\alpha \, \delta q_\alpha \Big|_{t_0}^{t_1} + \int_{t_0}^{t_1} \left[\delta p_\alpha \left(\dot{q}_\alpha - \frac{\partial \mathscr{H}}{\partial p_\alpha} \right) \right.$$
$$\left. - \delta q_\alpha \left(\dot{p}_\alpha + \frac{\partial \mathscr{H}}{\partial q_\alpha} \right) \right] dt = 0 \tag{10.5}$$

If we impose the condition that the varied path always passes through the given endpoints, the integrated part vanishes at the integration limits. At those points $\delta q_\alpha = 0$. Under the integral sign q_α and p_α are independent variables, the variations of which are absolutely arbitrary. Reasoning in exactly the same way as in

deriving the Lagrange equations in Section 2, we conclude that the variation δS can vanish only if the following equations are satisfied:

$$\dot{p}_\alpha = -\frac{\partial \mathcal{H}}{\partial q_\alpha} \tag{10.6a}$$

$$\dot{q}_\alpha = \frac{\partial \mathcal{H}}{\partial p_\alpha} \tag{10.6b}$$

Instead of n second-order Lagrange equations we obtain $2n$ first-order equations. They are called *Hamilton's equations.*

If $\mathcal{H}$ does not depend explicitly on time, the latter can be eliminated by dividing all the equations by one of them. For the sake of simplicity we shall show how this is done for a system with one degree of freedom:

$$\frac{dp}{dq} = -\frac{\partial \mathcal{H}}{\partial q} \bigg/ \frac{\partial \mathcal{H}}{\partial p} \tag{10.7}$$

Integration yields one arbitrary constant. The second constant is determined from the quadrature

$$\frac{dt}{dq} = \frac{1}{\partial \mathcal{H}/\partial p}$$

where $\mathcal{H}$ is a function of q and p in which p is expressed in terms of q with the help of the integrated equation (10.7). The integration constant of the last equation is simply the initial time t_0, which can be legitimately put equal to zero without loss of generality.

Note that we cannot substitute the momenta into the Lagrangian as we did into the energy to make use of all the known integrals of the motion (cf. (9.27)). Substitution of the momenta into the Lagrangian yields incorrect equations of motion for other variables. However, the following procedure can be adopted.

Let a certain generalized coordinate q_σ be cyclic. We express the corresponding (constant) momentum and solve the equation for it with respect to $\dot{q}_\sigma$. If we form the function

$$R = p_\sigma \dot{q}_\sigma - L \tag{10.8}$$

(there is no summation over σ), then with respect to q_σ it will be equivalent to the Hamiltonian, while remaining a Lagrangian for all the other variables. In other words, we must write Eqs. (10.6a) and (10.6b) for p_σ and q_σ, and the first one immediately yields $p_\sigma =$ $=$ constant. The equations of motion for all the other variables are formed with the help of R according to the rule (2.20). The function R is known as the *Routh function.*

Canonical Transformations. On many occasions we had to change the dynamic variables, in going over, for example, from orthogonal

to spherical coordinates. In such transformations the Lagrange equations do not, naturally, alter their form. They are written in the same way on the basis of the variation principle.

Now we must consider transformations of a more general type, affecting not only the generalized coordinates but the generalized momenta as well, insofar as both, as was shown, can be considered to be symmetrical with respect to one another. But there is one fundamental requirement for this: the transformed coordinates and momenta must satisfy requirements of the same form as the initial ones, that is (10.6a) and (10.6b), though, perhaps, with a different Hamiltonian. Hamilton's equations are sometimes called *canonical*, and the required transformations are therefore also known as canonical.

And so, assume that the old dynamic variables p_α and q_α are expressed in terms of new ones, P_β and Q_β, for which the Hamiltonian is not $\mathscr{H}$ but $\mathscr{K}$.

If the old Lagrangian was equal to $p_\alpha \dot{q}_\alpha - \mathscr{H}$, then the new one must have the form $P_\beta \dot{Q}_\beta - \mathscr{K}$. But we know from Section 2 that the Lagrangian is defined only up to the total derivative of some function of the coordinates and time. Let that function V be dependent on both the old and the new coordinates. Then the relationship between the two Lagrangians (the old and the new) is

$$p_\alpha \frac{dq_\alpha}{dt} - \mathscr{H} = P_\beta \frac{dQ_\beta}{dt} - \mathscr{K} + \frac{dV(q, Q)}{dt}$$

$$= P_\beta \frac{dQ_\beta}{dt} - \mathscr{K} + \frac{\partial V}{\partial q_\alpha} \frac{dq_\alpha}{dt} + \frac{\partial V}{\partial Q_\beta} \frac{dQ_\beta}{dt} + \frac{\partial V}{\partial t}$$

$$(10.9)$$

Since all dq_α's and dQ_β's are independent variables, Eq. (10.9) is valid only when their factors are equal. From this we obtain equations which must satisfy the old and new variables for the transformation from the one to the other to be canonical:

$$p_\alpha = \frac{\partial V}{\partial q_\alpha} \tag{10.10}$$

$$P_\beta = -\frac{\partial V}{\partial Q_\beta} \tag{10.11}$$

$$\mathscr{H} = \mathscr{K} - \frac{\partial V}{\partial t} \tag{10.12}$$

It was pointed out before that the coordinates and momenta are symmetrically involved in Hamilton's equations. We can therefore write transformation formulas which would contain, instead of old and new coordinates, old coordinates and new momenta, or old momenta and new coordinates, or old momenta and new momenta.

Let us show, for example, how to pass from transformations (10.10)-(10.12) to transformations containing q_α and P_β instead of q_α and Q_β. We put

$$V = V' - P_\beta Q_\beta, \quad V' = V'(q, P) \tag{10.13}$$

Then the first term in the right-hand side of (10.9), that is $P_\beta \, (dQ_\beta/dt)$, cancels out with the corresponding term from the transformation function, and we obtain the equation

$$p_\alpha \frac{dq_\alpha}{dt} - \mathscr{H} = -\mathscr{K} + \frac{\partial V'}{\partial q_\alpha} \frac{dq_\alpha}{dt}$$
$$+ \frac{\partial V'}{\partial P_\beta} \frac{dP_\beta}{dt} - Q_\beta \frac{dP_\beta}{dt} + \frac{\partial V'}{\partial t} \tag{10.14}$$

From it we find the other transformation formulas:

$$p_\alpha = \frac{\partial V'}{\partial q_\alpha} \tag{10.15}$$

$$Q_\beta = \frac{\partial V'}{\partial P_\beta} \tag{10.16}$$

$$\mathscr{H} = \mathscr{K} - \frac{\partial V'}{dt} \tag{10.17}$$

The method of obtaining the other formulas is obvious.

The Hamilton-Jacobi Equation. Suppose we have managed to find a transformation function $V\,(q, Q)$ such that the new Hamiltonian $\mathscr{K}$ is identically equal to zero. Then, from (10.6a) and (10.6b), the new dynamic variables satisfy the following equations:

$$\frac{dP_\beta}{dt} = -\frac{\partial \mathscr{K}}{\partial Q_\beta} = 0 \tag{10.18}$$

$$\frac{dQ_\beta}{dt} = \frac{\partial \mathscr{K}}{\partial P_\beta} = 0 \tag{10.19}$$

In other words, both P_β and Q_β are constant, and since solution of mechanical problems requires finding $2n$ constants, the function $V\,(q, Q)$, which makes the new Hamiltonian vanish, immediately yields the required solution. This function satisfies the equation

$$\mathscr{H}\left(q, \frac{\partial V}{\partial q}, t\right) = -\frac{\partial V}{\partial t} \tag{10.20}$$

This can be verified if we put $\mathscr{K} = 0$ in (10.12) and substitute the generalized momenta p from (10.15) into it.

The first-order partial differential equation (10.20) in n variables is called the *Hamilton-Jacobi equation*.

There is no need to seek a general solution of (10.20) containing $2n$ arbitrary functions. It is sufficient to find the so-called total integral

of the equation, which involves n arbitrary nonadditive constants C_i.

If such an integral has been found, the stated n constants can be taken as the new coordinates, that is, we can put

$$Q_i \equiv C_i$$

Then from (10.11) we obtain the new momenta:

$$P_i = -\frac{\partial V}{\partial Q_i} = -\frac{\partial V}{\partial C_i} \tag{10.21}$$

As can be seen from (10.18), these new momenta are also constant quantities. Thus, it is sufficient to find a solution of the Hamilton-Jacobi equation containing half of all the integrals of the motion. The rest are determined by simple differentiation.

Of course, the Hamilton-Jacobi equation can usually be solved in closed form only in those cases when the solution can be found in some other way. But the computations required to find the equations of motion are greatly simplified with the help of the Hamilton-Jacobi equation as compared with other methods of solving problems of mechanics. Besides, from the specific form of Eq. (10.20) it is easier to conclude whether it can have a solution in closed form.

Action as a Transformation Function. One possible transformation function is the action S of a system. To demonstrate this, consider Eq. (10.5), which expresses the total variation of the action. It was pointed out in Section 2 that the difference between the variation and the differential of a coordinate is that the former is arbitrary, while the differential is taken along the actual path. Let us now refer Eq. (10.5) to just that case, that is, actual motion. Then, since Hamilton's equations (10.6a) and (10.6b) are always satisfied for actual motion, the integrand vanishes. At the integration limits the variations δS become the differentials of those quantities taken along the paths, so that the action variation δS can be replaced by dS. We thus obtain

$$dS = p_\alpha \, dq_\alpha - p_\alpha^0 \, dq_\alpha^0 \tag{10.22}$$

But since the motion of the system has now been defined, the action S can be treated as a function of the running coordinates q_α for the given initial values q_α^0. This corresponds to the following construction. At the initial time we have a continuous set of systems of the same type differing only in the initial conditions of motion. In other words, the systems have the same Hamiltonians and at the initial time as it were fill a smooth surface so that each point of the surface corresponds to a certain set of initial conditions of motion.

With time each system develops its running values of the generalized coordinates in accordance with the dynamic laws and initial conditions. For these coordinates we can again construct a

surface similar to the initial one, but on it the value of the action will be different from what it was on the initial surface for the given system that "started" at time $t = 0$.

However, if we begin to "eject" systems from the initial surface in continuous succession, then at each instant there can be found a surface on which the value of the action is the same as it was at time
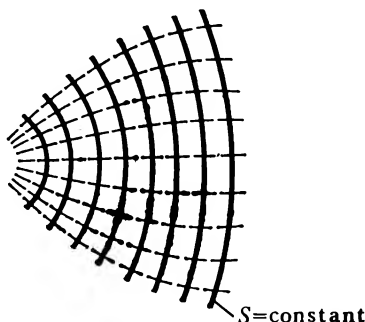

S=constant

Figure 13

$t = 0$ for systems "launched" at the beginning. Consequently, a constant-action surface moves in a q_α generalized-coordinate space without being fixed with respect to one and the same particles (we shall demonstrate this later with the simple example of free particles).

Consider the propagation of the surface of equal action shown schematically in Figure 13 in two-dimensional space. Since the action is now calculated along a known path, it is a function of q_α and t as well as of the point on the initial surface through which the given path passed. If $S = S(q_\alpha, q_\alpha^0, t)$, then for the given time t the total differential dS must be

$$dS = \left(\frac{\partial S}{\partial q_\alpha}\right)_{q^0} dq_\alpha + \left(\frac{\partial S}{\partial q_\alpha^0}\right)_q dq_\alpha^0 \tag{10.23}$$

From a comparison of (10.22) with (10.23) we conclude that

$$p_\alpha = \left(\frac{\partial S}{\partial q_\alpha}\right)_{q^0} \tag{10.24}$$

$$p_\alpha^0 = -\left(\frac{\partial S}{\partial q_\alpha^0}\right)_q \tag{10.25}$$

where the subscript of the parentheses indicates that all the variables from the given totality of q_α or q_α^0 are assumed constant in the differentiation.

Furthermore, we also calculate the partial derivative $\partial S/\partial t$. The total derivative dS/dt is equal to the Lagrangian (according to the definition of action). The partial derivative differs from the total derivative in the following way:

$$\frac{\partial S}{\partial t} = \frac{dS}{dt} - \frac{\partial S}{\partial q_\alpha} \frac{dq_\alpha}{dt}$$

Substituting p_α from (10.24), we obtain

$$\frac{\partial S}{\partial t} = L - p_\alpha \dot{q}_\alpha = -\mathscr{H} \tag{10.26}$$

But Eqs. (10.24)-(10.26) coincide with (10.15)-(10.17), if we put the new Hamiltonian $\mathscr{K}$ equal to zero in the latter. Hence, action is a transformation function which satisfies the Hamilton-Jacobi equation. The integration constants in this case are the initial coordinates, while the other n constants are, according to (10.21) and (10.25), the initial momenta.

Thus, the displacement of a system along a path of actual motion performs a continuously developing canonical transformation from the running variable coordinates and momenta to the constant initial coordinates and momenta.

Let us now find the speed with which the surfaces of constant action propagate through space. For clarity take a system with one degree of freedom. We differentiate the action and require the differential to vanish, since we want to find out how one and the same value of S propagates:

$$dS = \frac{\partial S}{\partial q} dq + \frac{\partial S}{\partial t} dt$$

Instead of the partial derivative of the action with respect to time we substitute the energy with a minus sign, and instead of $\partial S/\partial q$, the momentum. Hence the required velocity, dq/dt, is

$$\frac{dq}{dt} = \frac{E}{p_\zeta} \tag{10.27}$$

But from (10.6b) the velocity v of the particles themselves is equal not to the quotient E/p but to the derivative $\partial E/\partial p$. For example, in the case of free particles the surface of constant action propagates with the speed $mv^2/(2mv) = v/2$, that is, half as fast as the particles themselves.

Integration of the Hamilton-Jacobi Equation. As an example let us examine how the Hamilton-Jacobi equation is integrated in a specific case, Kepler's problem. As was shown in Section 3 (see (3.11)), the Lagrangian in this problem has the form

$$L = \frac{m}{2} (\dot{r}^2 + r^2 \dot{\vartheta}^2 + r^2 \sin^2 \vartheta \, \dot{\varphi}^2) - U(r)$$

Let us now find the Hamiltonian. The linear momenta are expressed in terms of the velocities as follows:

$$p_r = m\dot{r}, \qquad p_\vartheta = mr^2\dot{\vartheta}, \qquad p_\varphi = mr^2\sin^2\vartheta\,\dot{\varphi}$$

From this, expressing the generalized velocities in terms of the respective momenta and replacing in the Hamiltonian the quantity $-U$ involved in the Lagrangian by U, we obtain

$$\mathcal{H} = \frac{1}{2m}\left(p_r^2 + \frac{p_\vartheta^2}{r^2} + \frac{p_\varphi^2}{r^2\sin^2\vartheta}\right) + U(r) \qquad (10.28)$$

To obtain the Hamilton-Jacobi equation from this we replace the momenta by the derivatives of action with respect to the corresponding coordinate, according to the general rule, and equate the expression to $-\partial S/\partial t$. Instead of the transformation function V we write everywhere S. Then

$$\frac{1}{2m}\left[\left(\frac{\partial S}{\partial r}\right)^2 + \frac{1}{r^2}\left(\frac{\partial S}{\partial\vartheta}\right)^2 + \frac{1}{r^2\sin^2\vartheta}\left(\frac{\partial S}{\partial\varphi}\right)^2\right] + U(r) = -\frac{\partial S}{\partial t} \qquad (10.29)$$

First of all we must eliminate the variables not explicitly involved in the equation, that is, time t and azimuth φ. This is done as follows. Instead of S we introduce a new function S_0, the so-called "shortened" action:

$$S = -Et + p_\varphi\varphi + S_0(r, \vartheta) \qquad (10.30)$$

Substituting Eq. (10.30) into (10.29), we obtain the equation for S_0:

$$\frac{1}{2m}\left[\left(\frac{\partial S_0}{\partial r}\right)^2 + \frac{1}{r^2}\left(\frac{\partial S_0}{\partial\vartheta}\right)^2 + \frac{p_\varphi^2}{r^2\sin^2\vartheta}\right] + U(r) = E \qquad (10.31)$$

To integrate Eq. (10.31) we seek a solution in the form

$$S_0 = R(r) + \theta(\vartheta)$$

We substitute this solution into (10.31) and transpose the terms involving the variable ϑ to the left-hand side of the equation, and those involving energy to the right-hand side (though the latter is not essential). As a result, on the left we have a function dependent only on the polar angle. But r and ϑ are independent variables. We can, for example, vary r without changing ϑ. Then the left-hand side of the equation will change while the right-hand side remains unchanged, or vice versa. Such a situation may occur in one and only one case: when both sides of the equation are constant and do not depend on either r or ϑ. Denoting their value M^2, we obtain

two equations:

$$r^2 E - \frac{r^2}{2m}\left(\frac{dR}{dr}\right)^2 - U(r)\,r^2 = \frac{M^2}{2m}$$

$$\left(\frac{d\theta}{d\vartheta}\right)^2 + \frac{p_\varphi^2}{\sin^2\vartheta} = M^2$$

Integrating both with the help of quadratures and combining the result in Eq. (10.30), we obtain the total integral of the Hamilton-Jacobi equation, which involves three arbitrary constants:

$$S = -Et + p_\varphi\varphi + \int d\vartheta \left[M^2 - \frac{p_\varphi^2}{\sin^2\vartheta}\right]^{1/2}$$
$$+ \int dr \left[2m\left(E - U - \frac{M^2}{2mr^2}\right)\right]^{1/2} \quad (10.32)$$

From this we can determine other constants by differentiating with respect to the already known constants E, p_φ and M. The meaning of the first two new constants can be easily discerned from the answer: they are the initial time, and the initial azimuth taken with a minus sign. Indeed, after such substitutions the path equation is reduced to the form

$$t - t_0 = \int \frac{m\,dr}{[2m\,(E - U) - M^2/r^2]^{1/2}} \quad (10.33)$$

$$\varphi - \varphi_0 = \int \frac{p_\varphi\,d\vartheta}{\sin^2\vartheta\,(M^2 - p_\varphi^2/\sin^2\vartheta)^{1/2}} \quad (10.34)$$

$$\frac{\partial S}{\partial M} \equiv \alpha = \int \frac{M\,d\vartheta}{(M^2 - p_\varphi^2/\sin^2\vartheta)^{1/2}}$$
$$- \int \frac{M\,dr}{r^2\,[2m\,(E - U) - M^2/r^2]^{1/2}} \quad (10.35)$$

The integrals contained in (10.33) and (10.35) can be found by substituting the specific form of the dependence of the potential energy on the distance to the centre, $U = U(r)$. In practice the most important case is Kepler's problem, when $U = -a/r$. Then (10.33) expresses the time dependence of the radius r, (10.34) connects the azimuth and the polar angle, and (10.35) gives the dependence of the radius on the polar angle.

This type of solution is conveniently employed in cases involving several bodies moving about an attractive centre in different planes. Then the obtained solution holds as long as the interactions between the bodies are not taken into account. When perturbations are taken into account, such a solution yields a zero approximation.

Action Variables and Angular Variables. Let us now examine a solution of the Hamilton-Jacobi equation as applied to the problem

of finite motion of a system. An example of such a solution is (10.32) when the potential energy $U(r)$ refers to the forces of attraction, and the total energy $E < 0$ at $U(\infty) = 0$.

In finite motion in a central field r, ϑ and φ vary within finite limits, which are set for r and ϑ from the conditions that the radicands in (10.32) must be positive; φ varies from 0 to 2π. It can be seen from Eq. (10.24) that

$$p_r = [2m(E-U) - M^2/r^2]^{1/2}$$

$$p_\vartheta = (M^2 - p_\varphi^2/\sin^2\vartheta)^{1/2}$$

These momenta have real values for given E, M and p_φ when the radicands are positive.

Suppose the motion is finite over all variables. Then any of the integrals

$$J_i = \frac{1}{2\pi} \int p_i \, dq_i \tag{10.36}$$

converge over the whole domain of variations of p_i and q_i. The integration domain commences from any value of q_i, proceeds to the limit of the variation of q_i on one side (where the integrand vanishes), then back to the other limit, and to q_i again. In other words, it is the area of the curve $p_i = p_i(q_i)$ drawn in the plane in which p_i, q_i are laid off along the coordinate axes (2π is introduced into the definition so as to satisfy the equation

$$J_\varphi = \frac{1}{2\pi} \int\limits_0^{2\pi} p_\varphi \, d\varphi = p_\varphi$$

that is, so as for J_φ to simply coincide with p_φ; for uniformity the same factor is involved in all the J_k).

Suppose the integrals J_k have been calculated for all variables. Obviously, they depend only on the first integrals of the motion, which determine the values of all the generalized momenta p_α. For example, in the case of Kepler's problem J_r, J_ϑ, and J_φ are expressed in terms of E, M, p_φ. Let us now express the first integrals in terms of the quantities J_k. We obtain, in particular, the energy as a function of all the J_k's:

$$E = E(J_1, J_2, \ldots, J_k, \ldots, J_n) \tag{10.37}$$

We substitute the integrals of motion expressed in terms of J_k into an action function analogous to (10.32), but not involving time explicitly, that is, one into which the term $-Et$ has not been introduced. Then, according to the general theory of canonical transformations, we obtain a transformation function from the variables q_α to other, also canonical, variables. Let J_k be the gen-

eralized momenta in terms of the new variables. The transformation function now has the form

$$S = S_1(q_\alpha, J_k) \tag{10.38}$$

Since time is not involved explicitly in (10.38), the new Hamiltonian is, in accordance with (10.17), equal to the old one:

$$\mathcal{K} = \mathcal{H} = E(\dots J_k \dots) \tag{10.39}$$

But this new Hamiltonian depends only on the generalized momenta J_k, so that by applying Eqs. (10.6a) and (10.6b) we obtain Hamilton's equations for the momenta J_k and the corresponding coordinates, which we denote w_k:

$$\dot{J}_k = -\frac{\partial E}{\partial w_k} = 0 \tag{10.40}$$

$$\dot{w}_k = \frac{\partial E}{\partial J_k} = \text{constant} \tag{10.41}$$

The integration of (10.40) once again confirms that the J_k are constants of motion. From (10.41) it follows that the variables w_k vary linearly with time:

$$w_k = \frac{\partial E}{\partial J_k} t + w_k^0 \tag{10.42}$$

The quantities w_k and J_k are known as *angular variables* and *action variables*. The designation for w_k was chosen because these variables increase like an angle in uniform rotation.

Let us now follow the change of one angular variable due to the passing of the corresponding generalized coordinate over the whole permitted range of values in both directions. This is not actual motion, when all the generalized coordinates change; what we are doing is to fix the values of all the coordinates but one, which we vary for a given value of all the integrals of the motion $J_{i \neq k}$.

Obviously, when a varying coordinate passes through all possible values and returns, together with the corresponding velocity, to the initial value, only the angular variable w_k corresponding to it receives an increment. The action increment in this imaginary cycle is $2\pi J_k$, since over one cycle $\int p_i \, dq_i$ is equal to $2\pi J_i$ (see (10.36)). But according to (10.16), to such a variation of the action there must be correlated a change in the angular variable Δw_k, connected with it by the equation

$$\Delta w_k = \frac{\partial}{\partial J_k} \Delta S_k = \frac{\partial}{\partial J_k} 2\pi J_k = 2\pi$$

We find the time interval τ_k necessary for this cycle. From (10.42) we have

$$\Delta w_k = \frac{\partial E}{\partial J_k} \tau_k = 2\pi \tag{10.43}$$

To this period of change in the coordinate there corresponds a frequency ω_k equal, according to the general definition, to $2\pi/\tau_k$, that is

$$\omega_k = \frac{\partial E}{\partial J_k}$$

If only the coordinate q_k were changing, it would, bearing in mind the period of change τ_k, be dependent on time according to the law

$$q_k = \mathrm{Re}\,[\sum_{n=1,\,2,\,\ldots} A_{kn} \exp{(in\omega_k t)}] \tag{10.44}$$

Actually, though, all the q_k's depend on all the w_i's so that instead of the special law (10.44) we must write a more general one:

$$q_k = \mathrm{Re}\,[\sum_{n_1,\,n_2,\,\ldots,\,n_k} A_{kn_1 n_2\ldots n_k} \exp{(i \sum_h n_k \omega_k t)}] \tag{10.45}$$

where the integers n_k acquire any possible values.

Then, by "stopping" the time for all variables but one, we revert to (10.44), that is, to a periodic dependence, while in reality we have a dependence of more general form, Eq. (10.45).

If the frequencies ω_k are incommensurable, then all the linear combinations $\sum n_k \omega_k$ with integral n_k are different. Hence, for no value of t does (10.45) revert to the initial value it had at $t = 0$. If we wait long enough, however, it can come infinitely close to the initial value. The coordinate q_k is not periodic but is, as they say, an almost periodic, or quasi-periodic, function of time.

Sometimes, though, the frequencies are commensurable. In that case after a time interval that is a multiple of all periods the system reverts to its initial state, and its path closes. An example of a closed path is the elliptical path in Kepler's problem, in which all three periods τ_r, τ_ϑ, and τ_φ are the same. Another such case is a two-dimensional harmonic oscillator with equal oscillation frequencies in two directions.

Adiabatic Invariants. Suppose now that a system is subject to some external action dependent on time (for example, the Hamiltonian involves a parameter which gradually changes with time), but we accept that in any period τ_k the parameter λ varies but slightly.

If the Hamiltonian of the system involves time explicitly, the transformation function, S_1, to the variables J_k must also be dependent on time. Given these conditions, the old Hamiltonian is not equal to the new and is, according to (10.17), connected with it by

the relationship

$$\mathscr{K} = E\left(\ldots J_k \ldots\right) - \frac{\partial S_1}{\partial t} = E - \left(\frac{\partial S_1}{\partial \lambda}\right)_J \dot{\lambda} \qquad (10.46)$$

Here, at very slow variations of λ, the derivative $(\partial S_1/\partial \lambda)_J$ can be referred to a constant λ, and (10.46) can be treated as a Taylor expansion of the Hamiltonian involving only the first-order term in $\dot{\lambda}$.

If the Hamiltonian varies with time, its derivatives

$$\dot{J}_k = -\frac{\partial \mathscr{K}}{\partial w_k} = \left(\frac{\partial^2 S}{\partial w_k\, \partial \lambda}\right)_J \dot{\lambda} \qquad (10.47)$$

must also depend on time.

We shall show, however, that in a sense J_k varies much slower than λ. For this we introduce the concept of the mean of a certain quantity $f(t)$ with respect to time:

$$\bar{f} = \lim_{t \to \infty} \frac{1}{t} \int_0^t f\, dt \qquad (10.48)$$

Suppose that the time interval t is very great in comparison with all the periods τ_k of the system, but very small in comparison with the time of appreciable variation of the parameter λ. Or, if we write this condition in the form of strong inequalities: $\Delta\lambda = \dot{\lambda} t \ll \lambda$ and $\omega_k t \gg 1$, we can, in averaging Eq. (10.47) over time, take $\dot{\lambda}$ outside the integral. We obtain the following equation:

$$\bar{\dot{J}}_k = \frac{1}{t}\left[J_k(t) - J_k(0)\right] = \left(\frac{\partial^2 S}{\partial w_k\, \partial \lambda}\right)\dot{\lambda}$$

As can be seen from Eq. (10.46), the derivative $(\partial S/\partial \lambda)_J$ is taken at constant J_k. Therefore, if the time varies by any integral period τ_k (when to S is added $2\pi J_k$), the derivative receives no such increment, remaining a quasiperiodic function. But over a sufficiently long time interval the mean value of the function $\exp\left(i \sum_k n_k \omega_k t\right)$ tends to zero, because

$$\lim_{t \to \infty} \frac{1}{t} \int_0^t \exp\left(i \sum_k n_k \omega_k t\right) dt = \frac{\exp\left(i \sum_k n_k \omega_k t\right) - 1}{i \sum_k n_k \omega_k t} = 0$$

Consequently, the mean variation of the quantity $[J_k(t) - J_k(0)]/t$ is substantially less than the variation of λ in the same time. In the

limit, assuming $\dot{\lambda}$ to be an infinitesimal of the first order, we find that the variation of J_k is an infinitesimal of lower order. In other words, J_k should be considered as a constant quantity.

This is an important property of J_k that distinguishes it from energy, which varies accordingly as λ, or, in other words, is not conserved. The arbitrarily slow variation of a quantity in a system is known as *adiabatic variation*. Since J_k is conserved in such variation, it is called an *adiabatic invariant*.

Let us calculate it for a linear harmonic oscillator. According to the general definition (7.31), the energy of a separate oscillator is

$$E = \frac{P^2}{2} + \frac{\omega^2 Q^2}{2} \tag{10.49}$$

whence the adiabatic invariant is

$$J = \frac{2}{2\pi} \int_{-\sqrt{2E}/\omega}^{\sqrt{2E}/\omega} (2E - \omega^2 Q^2)^{1/2} dQ = \frac{2\pi E}{2\pi\omega} = \frac{E}{\omega} \tag{10.50}$$

Suppose that as a consequence of a variation in the elastic constant of an oscillator its frequency is slowly varying. Then we find from (10.50) that the energy must vary in proportion to the frequency.

If, for example, we gradually change the length of a pendulum, we can readily conclude with the help of (7.3) that the energy of the oscillations is inversely proportional to the square root of the length. Expressing energy in terms of the angular oscillation amplitude, φ_0, according to the formula $E = mgl\varphi_0^2/2$, we conclude that $\varphi_0 \propto l^{-3/4}$.

The conclusions concerning a pendulum can, of course, be obtained in elementary fashion. For this we must calculate the work against the tensile force of the thread done in shortening it over one oscillation period. If the thread was shortened by Δl, only one half the work done—the difference between the work done in raising the highest and lowest points of the pendulum—transforms into the oscillation energy. The tensile force in the thread due to the oscillations of the pendulum is equal to mv^2/l, where $v = l\omega\varphi_0 \sin \omega t$. Taking into account that the mean square of the sine in one oscillation is equal to $1/2$, we find that the work on extending the point of suspension that transforms into the oscillation energy is $-(1/2) mg\varphi_0^2\Delta l/2$. Since the oscillation energy itself is equal to $mgl\,\varphi_0^2/2$, we obtain the ratio

$$\frac{\Delta E}{E} = -\frac{\Delta l}{2l}$$

Passing from differences to differentials and integrating, we obtain

$$El^{1/2} = \text{constant}$$

EXERCISES

1. Find the transformation function from variables P, Q to variables J, w for a linear harmonic oscillator.

Solution. The action integral for a linear harmonic oscillator is

$$S_0' = \int P \, dQ = \int (2E - \omega^2 Q^2)^{1/2} \, dQ$$

According to the general rule, into this we must substitute the energy integral expressed in terms of the action variable J according to (10.50), or $E = J\omega$. Then the action

$$S_0' = J \arcsin Q \left(\frac{\omega}{2J}\right)^{1/2} + \frac{Q\omega}{2}\left(\frac{2J}{\omega} - Q^2\right)^{1/2}$$

In this form it is the transformation function involving the old coordinate Q and the new momentum J. From (10.13) and (10.16), the transformation function expressed in terms of the old coordinate Q and the new coordinate w is connected with S_0' by the relationship

$$S_0 = S_0^\bullet - wJ = S_0' - J\,\frac{\partial S_0'}{\partial J} = \frac{\omega Q}{2}\left(\frac{2J}{\omega} - Q^2\right)^{1/2}$$

This yields the required transformation function expressed in terms of the old coordinate and the new momentum. Into this we must substitute the new momentum, connected with the new coordinate by the formula

$$J = \frac{\omega Q^2}{2 \sin w}$$

We finally obtain the required function in the form

$$S_0 = \frac{\omega Q^2}{2} \cot w$$

(Poincaré's transformation).

2. Construct the equation of constant-action surfaces for a system of material particles emerging from one point in space with the same absolute velocity v_0 in a field of gravity.

Solution. Taking the initial point as the origin of the coordinate system, we have

$$v_x = v_{0x}, \qquad p_x = mv_{0x}, \qquad x = v_{0x}t$$
$$v_y = v_{0y}, \qquad p_y = mv_{0y}, \qquad y = v_{0y}t$$
$$v_z = v_{0z} - gt, \qquad p_z = mv_{0z} - mgt, \qquad z = v_{0z}t - gt^2/2$$

Eliminating the initial velocity conditions defining the path of each separate particle, we obtain for the surface of constant action S as a whole

$$p_x = m\,\frac{x}{t} = \frac{\partial S}{\partial x}, \qquad p_y = m\,\frac{y}{t} = \frac{\partial S}{\partial y}, \qquad p_z = \frac{mz}{t} - \frac{mgt}{2} = \frac{\partial S}{\partial z}$$

$$S = \frac{m}{2}\left(\frac{x^2}{t} + \frac{y^2}{t} + \frac{z^2}{t} - gzt - \frac{g^2 t^3}{12}\right)$$

From the expression for S we find the energy:

$$E = -\frac{\partial S}{\partial t} = \frac{m}{2}\left[\left(\frac{x}{t}\right)^2 + \left(\frac{y}{t}\right)^2 + \left(\frac{z}{t}\right)^2 + gz + \frac{g^2 t^2}{4}\right] = \frac{m v_0^2}{2}$$

3. Determine the action variables for Kepler's problem.

Solution. From (10.32) we determine the expressions for the **action** variables J_ϑ and J_r:

$$2\pi J_\vartheta = 2 \int_{\pi-\vartheta_0}^{\vartheta} (M^2 - p_\varphi^2/\sin^2\vartheta)^{1/2}\, d\vartheta$$

$$2\pi J_r = 2 \int_{r_1}^{r_2} [2m(E-U) - M^2/r^2]^{1/2}\, dr$$

where the radicands vanish at the integration limits. Substituting $U = -a/r$, we obtain after some simple computations

$$J_\vartheta = M - p_\varphi$$

$$J_r = |a|\left(\frac{m}{2|E|}\right)^{1/2} - M$$

Knowing, in addition, that $J_\varphi = p_\varphi$, we express the energy in terms of the action variables:

$$E = -\frac{a^2 m}{2(J_r + J_\vartheta + J_\varphi)^2}$$

Thus, all three rotation frequencies coincide:

$$\omega_r = \omega_\vartheta = \omega_\varphi = \frac{\partial E}{\partial J_r} = \frac{a^2 m}{(J_r + J_\vartheta + J_\varphi)^3}$$

which corresponds to a closed path.

4. The energy lost by the sun through radiation is generated at the expense of its mass. Determine how this affects the orbits of the planets.

Solution. We proceed from the equation of an orbit in its plane (5.12c). The constant a in the expression for the potential energy is equal to Gmm_0, where G is the gravitational constant. In the present case, the variable parameter λ is the mass m_0 of the sun. We express the eccentricity of the orbit ε, the major semiaxis b, and the rotation period τ in terms of the action variables, which do not change in a slow variation of the parameter m_0 (adiabatic invariants):

$$\varepsilon = \left(1 - \frac{J_\varphi^2}{(J_r + J_\varphi)^2}\right)^{1/2}, \qquad b = \frac{J_\varphi^2}{Gm^2 m_0(1-\varepsilon^2)}$$

$$\tau = \frac{2\pi}{\omega_r} = \frac{2\pi(J_r + J_\varphi)^3}{G^2 m^3 m_0^2}$$

PART II

ELECTRODYNAMICS

11

VECTOR ANALYSIS

The equations of electrodynamics are considerably simplified if they are written in vector, or in some cases tensor, form. Vector notation eliminates the arbitrariness associated with the choice of coordinate system and reveals the physical content of equations more vividly. In this sense tensor notation, which is employed in relativity theory, provides even greater advantages, making it possible to refer equations not only to arbitrary coordinates but to any inertial frame of reference.

In the foregoing discourse we assumed that the reader was familiar with elements of vector algebra, though we introduced certain clarifications. Furthermore, in Section 9 we offered a definition of tensors in which vector quantities represent a special case.

Now, in electrodynamics, vector differential operations are used. In the first part there was only one such operation: differentiation of a scalar with respect to a vector. This was used to calculate momentum from the Lagrangian or force from the potential energy formula. But it may be necessary to differentiate a vector with respect to a vector, and it is most useful to retain the vector notation, which offers a better idea of the geometric meaning of the operations.

This section is devoted to vector differential operations, as well as to tensor algebra, inasmuch as it is required in studying the theory of relativity.

Vector of an Area. We shall start with defining the vector of an area element $d\mathbf{S}$. This is a vector normal to the area, numerically equal to its surface, and related to the direction in which the bound-

ary of the area is traced as the displacement of a corkscrew is related to the direction of rotation of its handle (Figure 14). This definition covers, in particular, the *vector*, or *cross*, *product* of two vectors, which is numerically equal to the area of a parallelogram constructed with the two vectors as its sides, and, as a vector, is perpendicular to its plane. In Section 4 we offered detailed proof that the result
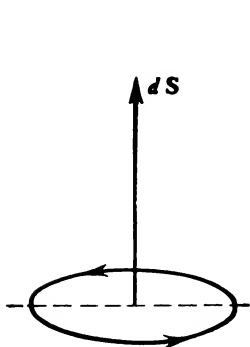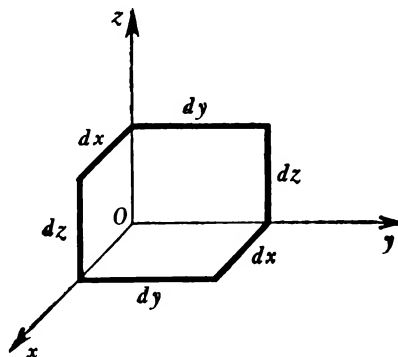

Figure 14   Figure 15

of a vector product is a vector, at least with respect to rotations of the coordinate system. This proof is also valid with respect to the vector of any area element.

We shall make use of a right-handed coordinate system x, y, z in which the rotation, looking from the z axis, is counterclockwise from x to y (Figure 15). In this coordinate system the area vector can be resolved into components as follows:

$$dS_x = dy\,dz, \qquad dS_y = dz\,dx, \qquad dS_z = dx\,dy$$

Flux of a Vector. Suppose now that a liquid of unit density ("water") is flowing across the area, the flow velocity being denoted by **v**. We denote the angle between dS and **v** as α. In Figure 16 are shown the streamlines passing through dS. They are parallel to the velocity **v**. Let us compute the per-second rate of flow of the liquid across the area dS. It is equal simply to $v\,dS'$, where dS' is the area perpendicular to the streamlines, as shown in Figure 16. Indeed, the amount of liquid passing through dS' in unit time is equal to a cylinder of base dS' and altitude v. But $dS' = dS\cos\alpha$, whence the required rate of flow of the liquid is

$$dJ = v\,dS' = v\,dS\cos\alpha = \mathbf{v}\,d\mathbf{S} \tag{11.1}$$

By analogy, the scalar product of any vector **A** (taken at the point of infinitesimal area) multiplied by $d\mathbf{S}$ is the *flux of vector* **A** across the area dS. Similar to the way that the flow of liquid across a finite area S is equal to the integral of dJ with respect to the area:

$$J = \int \mathbf{v} \, d\mathbf{S} \tag{11.2}$$

the integral

$$J = \int \mathbf{A} \, d\mathbf{S} \tag{11.3}$$

is called the flux vector **A** across the area.

The area vector is introduced so that we can make use   e convenient coordinate-free notation of (11.3). The integrals appearing
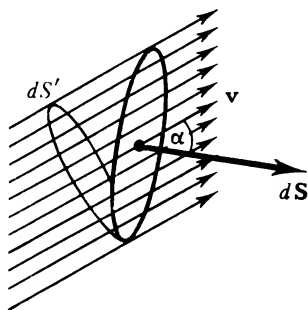
Figure 16

in (11.3) are double. In terms of the projections (11.3) can be writ en thus:

$$J = \int \mathbf{A} \, d\mathbf{S} \equiv \int\!\int A_x \, dy \, dz + \int\!\int A_y \, dz \, dx + \int\!\int A_z \, dy \, dx$$

where the limits of the double integrals are determined from the corresponding projections of the area boundary on the coordinate planes.

The Gauss Theorem. Let us now calculate the flux of a vector across a closed surface. For this we shall consider, first of all, the infinitesimal closed surface of a parallelepiped (Figure 17). We shall make the convention that the normal to a closed surface will always be taken outwards from the volume.

Let us calculate the flux of vector **A** across the area $ABCD$ (the direction of traverse being in agreement with the direction of the normal). Since the flux is equal to the scalar product of **A** by the

vector of area $ABCD$ in the negative x-direction (and hence equal to $-dy\,dz$), we obtain for this infinitely small area

$$dJ_{ABCD} = -A_x(x)\,dy\,dz$$

We get a similar expression for the area $A'B'C'D'$, only in this case the projection dS_x is equal to $dy\,dz$, and A_x is taken at the
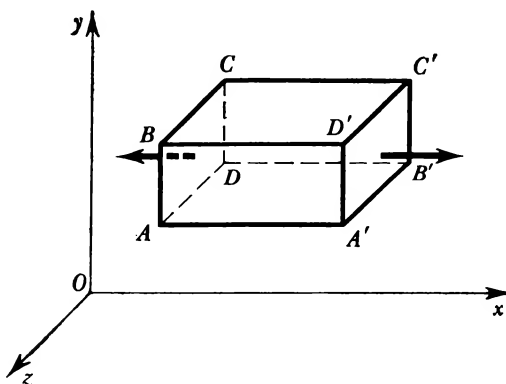

Figure 17

point $x + dx$ instead of x. Therefore

$$dJ_{A'B'C'D'} = A_x(x+dx)\,dy\,dz$$

Thus, the resultant flux across both areas perpendicular to the x axis is

$$dJ_{A'B'C'D'} + dJ_{ABCD} = [A_x(x+dx) - A_x(x)]\,dy\,dz$$

$$= \frac{\partial A_x}{\partial x}\,dx\,dy\,dz \tag{11.4}$$

We have utilized the fact that dx is an infinitely small quantity, and we have expanded $A_x(x + dx)$ in a series $A_x(x + dx) = A_x(x) + (\partial A_x/\partial x)\,dx$. The resultant fluxes across the boundaries perpendicular to the y and z axes are formed similarly. The net flux across the whole parallelepiped is

$$dJ = \left(\frac{\partial A_x}{\partial x} + \frac{\partial A_y}{\partial y} + \frac{\partial A_z}{\partial z}\right) dx\,dy\,dz \tag{11.5}$$

A finite closed volume can be divided into small parallelepipeds, and the relationship (11.5) applied to each one of them separately. If we sum all the fluxes, the adjacent boundaries do not contribute, since the flux emerging from one parallelepiped enters the neighbouring one. Only the fluxes through the outer surface of the selected

volume remain, since they are not cancelled by others. But the right-hand sides of (11.5) will be additive for all the volume elements $dV = dx\, dy\, dz$, yielding the very important integral theorem:

$$\int \mathbf{A}\, d\mathbf{S} = \int \left(\frac{\partial A_x}{\partial x} + \frac{\partial A_y}{\partial y} + \frac{\partial A_z}{\partial z} \right) dV \qquad (11.6)$$

It is called the *Gauss theorem.*

The Divergence of a Vector. The expression appearing in the right-hand side under the integral sign can be written in a much shorter form. We first of all note that it is a scalar expression, since there is a scalar in the left-hand side of (11.6) (the "mass of water"), and dV is also a scalar. This expression is called the *divergence of vector* $\mathbf{A}$ and is written thus:

$$\operatorname{div} \mathbf{A} \equiv \frac{\partial A_x}{\partial x} + \frac{\partial A_y}{\partial y} + \frac{\partial A_z}{\partial z} \qquad (11.7)$$

Divergence can be defined independently from any coordinate system, if (11.5) is used. Indeed, from (11.5) the difinition for divergence follows as

$$\operatorname{div} \mathbf{A} = \lim_{V \to 0} \frac{\int \mathbf{A}\, d\mathbf{S}}{V} \qquad (11.8)$$

The divergence of a vector at a given point is equal to the limit of the ratio of the vector flux across the surface surrounding the point to the volume enveloped by the surface, when the surface is contracted to a point.

Let us suppose that $\mathbf{A}$ denotes the velocity field of some fluid. Then, from definition (11.8), it can be seen that the divergence of $\mathbf{A}$ is a measure of the density of the sources of the fluid, or the number of sources per unit volume from which a unit mass of the fluid flows in unit time. Obviously, the more sources there are per unit volume, the more fluid will flow out of it. If $\operatorname{div} \mathbf{A}$ is negative, we can speak of the density of sinks. But it is more convenient to define the source density with the corresponding sign. We note that from (11.7) there follows the quantity

$$\operatorname{div} \mathbf{r} = \frac{\partial x}{\partial x} + \frac{\partial y}{\partial y} + \frac{\partial z}{\partial z} = 1 + 1 + 1 = 3 \qquad (11.9)$$

since $\mathbf{r}$ has components x, y, z.

It is not hard to derive the same result from the definition of divergence (11.8) not involving the coordinates. First, from the origin of the coordinate system we construct a cone containing an infinitesimal solid angle at the z vertex (Sec. 6). Since the radius vector coincides with the generatrix and is consequently perpendicu-

lar to the normal to the element of the side surface of the cone, there is no vector flux across the surface: the vector slips along without penetrating it. The radius vector flux only crosses the base of the cone.

Since the cone angle is infinitesimal, the base is equal to $r^2 \, d\Omega$. The radius vector is perpendicular to the base at all points, hence the flux of the radius vector across the base is $r \times r^2 \, d\Omega$. The volume of a cone is equal to the product of the base area times one-third the altitude, that is, again the radius. Substituting the flux of the radius vector and the volume of the cone into (11.8), we obtain div $\mathbf{r} = 3$. Actually in this proof there was no need to assume the cone infinitely small. We did this only to make use of the general definition of the divergence of a vector.

Circulatory Integrals and Stokes' Theorem. Let us consider the vector integral over a closed contour:

$$C = \int \mathbf{A} \, d\mathbf{l} = \int (A_x \, dx + A_y \, dy + A_z \, dz) \qquad (11.10)$$

This single integral is called the *circulation of the vector* over the closed contour. For example, if $\mathbf{A}$ is the force acting on any particle,
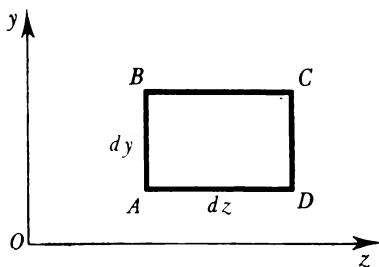

Figure 18

then $\mathbf{A} \, d\mathbf{l} = A \, dl \cos \alpha$ is the work done by the force on the contour element $d\mathbf{l}$, and C is the work done in covering the whole contour.

Now let us prove that the circulation of a vector $\mathbf{A}$ around a contour can be replaced by the integral over the surface stretched on the contour. Consider the projection of an infinitely small rectangular contour on the y,z-plane. Let this projection also have the form of the rectangle in Figure 18. We shall calculate the circulation of $\mathbf{A}$ around this rectangle. The side AB contributes a component $A_y(z) \, dy$, and side CD a component $-A_y(z + dz) \, dy$, where the minus sign must be written because the direction of the vector CD is opposite to that of the vector AB. We obtain, for the sum due

to the sides AB and CD,

$$-A_y(z+dz)\,dy + A_y(z)\,dy = -\frac{\partial A_y}{\partial z}\,dy\,dz$$

(we have expanded $A_y(z+dz)$ in a series in dz). For the sides BC and DA,

$$A_z(y+dy)\,dz - A_z(y)\,dz = \frac{\partial A_z}{\partial y}\,dy\,dz$$

The resultant value for circulation in the y,z-plane is

$$dC_{yz} = \left(\frac{\partial A_z}{\partial y} - \frac{\partial A_y}{\partial z}\right)dy\,dz = \left(\frac{\partial A_z}{\partial y} - \frac{\partial A_y}{\partial z}\right)dS_x \equiv B_x\,dS_x \tag{11.11}$$

The notation B_x is clear from the equation.

If the contour is arbitrarily oriented in space, all its infinitesimal segments must, in accordance with (11.10), be projected on the three coordinate axes. Then the whole contour has projections on the three coordinate planes, and the circulation resolves into the sum of three expressions of the form (11.11).

The circulation taken along the small contour is

$$dC = B_x\,dS_x + B_y\,dS_y + B_z\,dS_z = \mathbf{B}\,d\mathbf{S} \tag{11.12}$$

where B_x, B_y, and B_z are the abbreviated notations of the following differences between partial derivatives:

$$B_x = \frac{\partial A_z}{\partial y} - \frac{\partial A_y}{\partial z} \tag{11.13}$$

$$B_y = \frac{\partial A_x}{\partial z} - \frac{\partial A_z}{\partial x} \tag{11.14}$$

$$B_z = \frac{\partial A_y}{\partial x} - \frac{\partial A_x}{\partial y} \tag{11.15}$$

Circulation is, by definition (11.10), a scalar quantity. Consequently, the quantity in the right-hand side of (11.12) is also a scalar. But since $d\mathbf{S}$ is a vector, $\mathbf{B}$ is also a vector whose components are defined by Eqs. (11.13)-(11.15).

Vector $\mathbf{B}$ has a special name, the *rotation*, or *curl*, *of vector* $\mathbf{A}$, and is denoted thus:

$$\mathbf{B} = \operatorname{curl} \mathbf{A}$$

Curl $\mathbf{A}$ is resolved along the coordinate axes with the help of three unit vectors:

$$\mathbf{B} = \operatorname{curl} \mathbf{A}$$

$$= \mathbf{n}^{(x)}\left(\frac{\partial A_z}{\partial y} - \frac{\partial A_y}{\partial z}\right) + \mathbf{n}^{(y)}\left(\frac{\partial A_x}{\partial z} - \frac{\partial A_z}{\partial x}\right) + \mathbf{n}^{(z)}\left(\frac{\partial A_y}{\partial x} - \frac{\partial A_x}{\partial y}\right) \tag{11.16}$$

Comparing (11.16) with (11.12), we see that Eq. (11.12) contains the component of curl **A** normal to the area:

$$\int \mathbf{A} \, d\mathbf{l} = \mathrm{curl}_n \, \mathbf{A} \, dS \tag{11.17}$$

where the subscript n of curl **A** indicates that we should take not the whole vector curl **A** but only its projection on the normal to the area, dS. Equation (11.17) enables us to determine curl **A** in a coordinate-free manner, similar to the way we defined div **A** in (11.8), namely

$$\mathrm{curl}_n \, \mathbf{A} = \lim_{S \to 0} \frac{\int \mathbf{A} \, d\mathbf{l}}{S} \tag{11.18}$$

so that the normal component of curl **A** to any area at a given point in space is the limit of the ratio of the circulation along the boundary of the area to its surface, when the boundary contracts into a point.

For the integral $\int \mathbf{A} \, d\mathbf{l}$ to be nonzero, we must have closed vector lines, to some extent following the integration line, which lines are similar to the closed lines of flow in a liquid in vortex motion. Hence the term curl, or rotation.

If the circulation is calculated for a finite contour, then the contour can be broken up into infinitely small cells to form a grid. For the sides of adjacent cells, the circulations mutually cancel since each side is traversed twice in opposite directions; only the circulation along the external contour itself remains. The integral in the right-hand side of Eq. (11.17) gives the flux of curl **A** across the surface stretched on the contour. Thus, we obtain the desired integral theorem

$$\int \mathbf{A} \, d\mathbf{l} = \int \mathrm{curl} \, \mathbf{A} \, d\mathbf{S} \tag{11.19}$$

which is called *Stokes' theorem*.

Differentiation with Respect to the Radius Vector. The divergence and curl of a vector are its derivatives with respect to the vector argument. They can be reduced to a unified notation in the following way. We introduce the vector symbol ∇ (del, or nabla[1]) with components

$$\nabla_x \equiv \frac{\partial}{\partial x}, \qquad \nabla_y \equiv \frac{\partial}{\partial y}, \qquad \nabla_z \equiv \frac{\partial}{\partial z} \tag{11.20}$$

Then premultiplication by del denotes differentiation with respect to the radius vector. But there are two ways of multiplying vectors.

---

[1] After an ancient harp whose form ∇ resembles.

The scalar product of the vector by vector **A** should look like this:

$$(\nabla \cdot \mathbf{A}) \equiv \nabla_x A_x + \nabla_y A_y + \nabla_z \mathbf{A}_z = \frac{\partial A_x}{\partial x} + \frac{\partial A_y}{\partial y} + \frac{\partial A_z^{\mathfrak{k}}}{\partial z} \equiv \operatorname{div} \mathbf{A}$$

(11.21)

Another method is vector multiplication, which is defined as follows:

$$\nabla \times \mathbf{A} \equiv \mathbf{n}^{(x)}(\nabla_y A_z - \nabla_z A_y) + \mathbf{n}^{(y)}(\nabla_z A_x - \nabla_x A_z)$$
$$+ \mathbf{n}^{(z)}(\nabla_x A_y - \nabla_y A_x) \equiv \operatorname{curl} \mathbf{A} \quad (11.22)$$

We have used the identity sign everywhere to emphasize that we are simply dealing with a new system of notation. It is very convenient in vector analysis, because the operations are graphic and the equations are concise. Use of the del operator in the proof of various general relationships makes it simply unnecessary to resolve vectors into components.

In algebraic operations the del is in every way similar to a conventional vector. Multiplication by del signifies its operation on the given expression, if it is differentiated. Sometimes the del is multiplied by a vector (usually post-multiplied) without operating on it as a derivative. In that case it operates on another vector (see (11.30) and (11.32)).

If we operate with ∇ on a scalar φ, we obtain a vector which is called the *gradient of scalar* φ:

$$\operatorname{grad} \varphi \equiv \nabla_q \equiv \mathbf{n}^{(x)} \frac{\partial \varphi}{\partial x} + \mathbf{n}^{(y)} \frac{\partial \varphi}{\partial y} + \mathbf{n}^{(z)} \frac{\partial \varphi}{\partial z} \quad (11.23)$$

Its projections are:

$$\nabla_x \varphi = \frac{\partial \varphi}{\partial x}, \qquad \nabla_y \varphi = \frac{\partial \varphi}{\partial y}, \qquad \nabla_z \varphi = \frac{\partial \varphi}{\partial z} \quad (11.24)$$

From Eqs. (11.24), it can be seen that the vector $\nabla \varphi$ is perpendicular to the surface $\varphi = $ constant. Indeed, if we take a vector $d\mathbf{l}$ lying on this surface, then, in a displacement $d\mathbf{l}$, φ does not change (by the definition of $d\mathbf{l}$). This is written as

$$d\varphi = \frac{\lceil \partial \varphi}{\partial x} dl_x + \frac{\partial \varphi}{\partial y} dl_y + \frac{\partial \varphi^{\rceil}}{\partial z} dl_z = (\nabla \varphi \cdot d\mathbf{l}) = 0 \quad (11.25)$$

that is, $\nabla \varphi$ is perpendicular to any vector which lies in the plane tangential to the surface $\varphi = $ constant at the given point, which accords with our assertion.

Differentiation of Products. In considering the rules of differential operations with ∇ it should be remembered that with respect to differentiation the del is a derivative sign, while with respect to the

rules of transformation of a coordinate system it is a vector that is multiplied like any other vector.

First of all, the gradient of the product of two scalars is calculated as the derivative of a product:

$$\text{grad}\,(\varphi\psi) = \nabla\varphi\psi = \varphi\,\nabla\psi + \psi\,\nabla\varphi$$

$$= \psi\,\text{grad}\,\varphi + \varphi\,\text{grad}\,\psi \tag{11.26}$$

The divergence of a product of a scalar with a vector is calculated thus:

$$\text{div}\,\varphi\mathbf{A} = (\nabla_\varphi \cdot \varphi\mathbf{A}) + (\nabla_\mathbf{A} \cdot \varphi\mathbf{A}) = \mathbf{A}\,(\nabla\varphi) + \varphi\,(\nabla\cdot\mathbf{A})$$

$$= \mathbf{A}\,\text{grad}\,\varphi + \varphi\,\text{div}\,\mathbf{A} \tag{11.27}$$

Here the indices φ and $\mathbf{A}$ attached to ∇ show what ∇ is applied to.

We find the curl of $\varphi\mathbf{A}$ in a similar manner:

$$\text{curl}\,\varphi\mathbf{A} = \nabla_\varphi \times \varphi\mathbf{A} + \nabla_\mathbf{A} \times \varphi\mathbf{A}$$

$$= \text{grad}\,\varphi \times \mathbf{A} + \varphi\,\text{curl}\,\mathbf{A} \tag{11.28}$$

Now we shall operate with ∇ on the product of two vectors:

$$\text{div}\,(\mathbf{A} \times \mathbf{B}) = \nabla\,(\mathbf{A} \times \mathbf{B})$$

$$= \nabla_\mathbf{A}^!\,(\mathbf{A} \times \mathbf{B}) + \nabla_\mathbf{B}\,(\mathbf{A} \times \mathbf{B})$$

We perform a cyclic permutation in both terms, since ∇ can be treated in the same way as an ordinary vector. In addition, we have put $\mathbf{B}$ after $\nabla_\mathbf{B}$ in the second term, and here, as usual, we must change the sign of the vector product. The result is

$$\text{div}\,(\mathbf{A} \times \mathbf{B}) = \mathbf{B}\,(\nabla_\mathbf{A} \times \mathbf{A}) - \mathbf{A}\,(\nabla_\mathbf{B} \times \mathbf{B})$$

$$= \mathbf{B}\,\text{curl}\,\mathbf{A} - \mathbf{A}\,\text{curl}\,\mathbf{B} \tag{11.29}$$

Let us find the curl of a vector product. Here we must use the relationship $\mathbf{A} \times (\mathbf{B} \times \mathbf{C}) = \mathbf{B}\,(\mathbf{A}\cdot\mathbf{C}) - \mathbf{C}\,(\mathbf{A}\cdot\mathbf{B})$ to get

$$\text{curl}\,(\mathbf{A} \times \mathbf{B}) = \nabla\times(\mathbf{A}\times\mathbf{B}) = \nabla_\mathbf{A} \times(\mathbf{A}\times\mathbf{B}) + \nabla_\mathbf{B} \times(\mathbf{A}\times\mathbf{B})$$

$$= (\mathbf{B}\cdot\nabla_\mathbf{A})\mathbf{A} - (\nabla_\mathbf{A}\cdot\mathbf{A})\mathbf{B} + (\nabla_\mathbf{B}\cdot\mathbf{B})\mathbf{A} - (\mathbf{A}\cdot\nabla_\mathbf{B})\mathbf{B}$$

$$= (\mathbf{B}\cdot\nabla)\mathbf{A} - \mathbf{B}\,\text{div}\,\mathbf{A} + \mathbf{A}\,\text{div}\,\mathbf{B} - (\mathbf{A}\cdot\nabla)\,\mathbf{B}$$

$$\tag{11.30}$$

We note the new symbols $(\mathbf{B}\cdot\nabla)$ and $(\mathbf{A}\cdot\nabla)$ operating on the vectors $\mathbf{A}$ and $\mathbf{B}$. Obviously, $(\mathbf{A}\cdot\nabla)$ and $(\mathbf{B}\cdot\nabla)$ are symbolic scalars, equal, by definition of ∇, to

$$(\mathbf{A}\cdot\nabla) \equiv A_x\nabla_x + A_y\nabla_y + A_z\nabla_z \equiv A_x\frac{\partial}{\partial x} + A_y\frac{\partial}{\partial y} + A_z\frac{\partial}{\partial z}$$

$$\tag{11.31}$$

and similarly for $(\mathbf{B} \cdot \nabla)$. Then, $(\mathbf{A} \cdot \nabla) \mathbf{B}$ is a vector which is obtained by application of the scalar operation $(\mathbf{A} \cdot \nabla)$ to all the components of $\mathbf{B}$, in accordance with (11.31). Since $(\mathbf{A} \cdot \nabla)$ is a scalar operation it can be introduced under the vector multiplication sign—dot or cross—keeping in mind its differential properties.

Now we only have to compute the gradient of a scalar product with operations of this type:

$$\nabla (\mathbf{A} \cdot \mathbf{B}) = \nabla_{\mathbf{A}} (\mathbf{A} \cdot \mathbf{B}) + \nabla_{\mathbf{B}} (\mathbf{A} \cdot \mathbf{B})$$

We use the same transformation as in the preceding case:

$$\begin{aligned}
\text{grad} (\mathbf{A} \cdot \mathbf{B}) = (\mathbf{B} \cdot \nabla_{\mathbf{A}}) \mathbf{A} &+ \mathbf{B} \times (\nabla_{\mathbf{A}} \times \mathbf{A}) \\
&+ (\mathbf{A} \cdot \nabla_{\mathbf{B}}) \mathbf{B} + \mathbf{A} \times (\nabla_{\mathbf{B}}^{\square} \times \mathbf{B}) \\
= (\mathbf{B} \cdot \nabla) \mathbf{A} &+ \mathbf{B} \times \text{curl} \, \mathbf{A} \\
&+ (\mathbf{A} \cdot \nabla) \mathbf{B} + \mathbf{A} \times \text{curl} \, \mathbf{B} \quad (11.32)
\end{aligned}$$

Certain Special Formulas. We note certain essential cases of operations involving ∇.

From the definition of divergence (11.7), we obtain from (11.27) and (11.9)

$$\text{div} \frac{\mathbf{r}}{r^3} = \frac{1}{r^3} \text{div} \, \mathbf{r} + \mathbf{r} \, \text{grad} \, \frac{1}{r^3} = \frac{3}{r^3} - \frac{3 (\mathbf{r} \cdot \mathbf{r})}{r^5} = 0 \qquad (11.33)$$

In taking the gradient of r^{-3} we applied the rule for differentiating a composite function: we first differentiated r^{-3} with respect to r and then took ∇r. This is done as follows. Knowing that $r = = (x^2 + y^2 + z^2)^{1/2}$, we find

$$\frac{\partial r}{\partial x} = \nabla_x r = \frac{x}{(x^2 + y^2 + z^2)^{1/2}} = \frac{x}{r}$$

Going over from the component with respect to x to vector notation, we obtain

$$\text{grad} \, r = \nabla r = \frac{\mathbf{r}}{r} \qquad (11.34)$$

whence

$$\nabla \frac{1}{r^3} = \frac{d}{dr} \left(\frac{1}{r^3} \right) \nabla r = - \frac{3}{r^4} \frac{\mathbf{r}}{r}$$

The curl of a radius vector is zero. For example, for the component along the x axis we have

$$\text{curl}_x \, \mathbf{r} = \frac{\partial z}{\partial y} - \frac{\partial y}{\partial z} = 0$$

and in general

$$\text{curl} \, \mathbf{r} = 0 \qquad (11.35)$$

We now take

$$(\mathbf{A}\cdot\nabla)\, x = A_x \frac{\partial x}{\partial x} + A_y \frac{\partial x}{\partial y} + A_z \frac{\partial x}{\partial z} = A_x$$

and for all components of **r**

$$(\mathbf{A}\cdot\nabla)\, \mathbf{r} = \mathbf{A} \tag{11.36}$$

We shall now show how to operate with ∇ on a vector whose components depend only upon the absolute value of the radius vector **r**. As in the case of a scalar quantity, the rule of differentiating a composite function must be applied, with $\mathbf{r}/r$ substituted for ∇r according to (11.34). For the divergence of **A** (r) we obtain

$$\operatorname{div}\mathbf{A}\,(r) = \left(\frac{d\mathbf{A}}{dr}\cdot\nabla r\right) = \frac{1}{r}\,(\dot{\mathbf{A}}\cdot\mathbf{r}) \tag{11.37}$$

where $\dot{\mathbf{A}}$ is a total derivative of **A** (r) with respect to the argument r, that is, a vector whose components are the derivatives of the three components of **A** (r) with respect to r: $\dot{A}_x$, $\dot{A}_y$, $\dot{A}_z$. Further,

$$\operatorname{curl}\mathbf{A}\,(r) = \nabla r \times \dot{\mathbf{A}} = \frac{1}{r}\,(\mathbf{r}\times\dot{\mathbf{A}}) \tag{11.38}$$

Repeated Differentiation. Let us investigate certain results concerning repeated operations with ∇.

The curl of the gradient of a scalar is equal to zero:

$$\operatorname{curl}\operatorname{grad}\varphi = \nabla \times \nabla\varphi = (\nabla \times \nabla)\,\varphi = 0 \tag{11.39}$$

since the vector product of any vector (including ∇) by itself is equal to zero. This can also be seen by expanding curl grad φ in terms of its components. The divergence of a curl is also equal to zero:

$$\operatorname{div}\operatorname{curl}\mathbf{A} = \nabla\,(\nabla \times \mathbf{A}) = (\nabla \times \nabla)\,\mathbf{A} = 0 \tag{11.40}$$

Let us write down the divergence of the gradient of a scalar φ in component form. From Eqs. (11.7) and (11.24) we have

$$\operatorname{div}\operatorname{grad}\varphi = (\nabla\cdot\nabla)\,\varphi = \frac{\partial^2\varphi}{\partial x^2} + \frac{\partial^2\varphi}{\partial y^2} + \frac{\partial^2\varphi}{\partial z^2} \equiv \nabla^2\varphi \tag{11.41}$$

Here ∇^2 is the so-called *Laplace operator*, or *Laplacian*:

$$\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}$$

Finally, the curl of a curl can be expanded as a double vector product:

$$\operatorname{curl}\operatorname{curl}\mathbf{A} = \nabla \times (\nabla \times \mathbf{A}) = \nabla\,(\nabla\cdot\mathbf{A}) - (\nabla\cdot\nabla)\,\mathbf{A}$$
$$= \operatorname{grad}\operatorname{div}\mathbf{A} - \nabla^2\mathbf{A} \tag{11.42}$$

The latter equation is usually used to determine $\nabla^2 \mathbf{A}$, since in curvilinear coordinates $\nabla^2 \varphi$ and $\nabla^2 \mathbf{A}$ are expressed differently. It cannot be said that $\nabla^2 \mathbf{A}$ is the divergence of a vector gradient, because the gradient of a vector has not been defined.

Curvilinear Coordinates. In many problems it is useful to go over from rectilinear to curvilinear coordinates. We shall now show how various vector operations can be written down in curvilinear coordinates.

Curvilinear coordinates q_1, q_2, q_3 are termed orthogonal if only the quadratic terms dq_1^2, dq_2^2, dq_3^2 appear in the expression for the
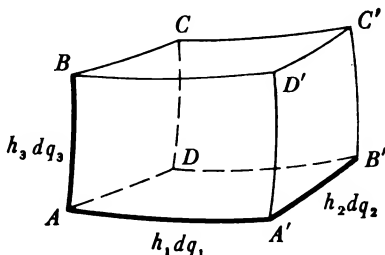
Figure 19

element of length dl^2, and not the products $dq_1 dq_2$, $dq_1 dq_3$, $dq_2 dq_3$, similar to the way that $dl^2 = dx^2 + dy^2 + dz^2$ appears in rectangular coordinates. In orthogonal coordinates

$$dl^2 = h_1^2\, dq_1^2 + h_2^2\, dq_2^2 + h_3^2\, dq_3^2 \qquad (11.43)$$

For example, in spherical coordinates $q_1 = r$, $q_2 = \vartheta$, $q_3 = \varphi$.
The element of length in spherical coordinates is (see (3.10))

$$dl^2 = dr^2 + r^2\, d\vartheta^2 + r^2 \sin^2 \vartheta\, d\varphi^2$$

so that

$$h_1 = 1, \quad h_2 = r, \quad h_3 = r \sin \vartheta$$

Let us construct an elementary parallelepiped (Figure 19). Then the components of the gradient are:

$$\mathrm{grad}_1 \psi = \frac{1}{h_1} \frac{\partial \psi}{\partial q_1}$$

$$\mathrm{grad}_2 \psi = \frac{1}{h_2} \frac{\partial \psi}{\partial q_2}$$

$$\mathrm{grad}_3 \psi = \frac{1}{h_3} \frac{\partial \psi}{\partial q_3} \qquad (11.44)$$

In order to find the divergence we repeat the proof of the Gauss theorem for Figure 19. The area $ADCB$ is equal to $h_2 h_3\, dq_2\, dq_3$. The flux of vector $\mathbf{A}$ across it is

$$A_1(q_1)\, h_2 h_3\, dq_2\, dq_3$$

Here, h_2 and h_3 are taken for the same value of q_1 as A_1 is. The sum of the fluxes across the areas $ADCB$ and $A'B'C'D'$ is

$$\frac{\partial}{\partial q_1}\, (h_2 h_3 A_1)\, dq_1\, dq_2\, dq_3$$

where we have used the expansion of the quantity $h_2 h_3 A_1$ at the point $q_1 + dq_1$ in terms of dq_1, in a way similar to (11.4). The total flux across all the boundaries is

$$dJ = \left[\, \frac{\partial}{\partial q_1}\, (h_2 h_3 A_1) + \frac{\partial}{\partial q_2}\, (h_3 h_1 A_2) \right.$$
$$\left. + \frac{\partial}{\partial q_3}\, (h_1 h_2 A_3) \right] dq_1\, dq_2\, dq_3$$

Let us now take advantage of the definition of divergence (11.8) to get

$$dJ = \operatorname{div} \mathbf{A}\, dV = \operatorname{div} \mathbf{A}\, (h_1 h_2 h_3\, dq_1\, dq_2\, dq_3)$$

Hence

$$\operatorname{div} \mathbf{A} = \frac{1}{h_1 h_2 h_3} \left[\, \frac{\partial}{\partial q_1}\, (h_2 h_3 A_1) + \frac{\partial}{\partial q_2}\, (h_3 h_1 A_2) \right.$$
$$\left. + \frac{\partial}{\partial q_3}\, (h_1 h_2 A_3) \right] \quad (11.45)$$

If instead of A_1, A_2, A_3 we substitute the expressions (11.44), the result will be the Laplacian of a scalar in orthogonal curvilinear coordinates. In spherical coordinates it is

$$\nabla^2 \psi = \frac{1}{r^2}\, \frac{\partial}{\partial r}\, r^2\, \frac{\partial \psi}{\partial r} + \frac{1}{r^2 \sin \vartheta}\, \frac{\partial}{\partial \vartheta}\, \sin \vartheta\, \frac{\partial \psi}{\partial \vartheta}$$
$$+ \frac{1}{\sin^2 \vartheta\, r^2}\, \frac{\partial^2 \psi}{\partial \varphi^2} \quad (11.46)$$

With the aid of Stokes' theorem, we can also calculate the curl in curvilinear coordinates. Without repeating the proof of theorem (11.19), we write:

$$\operatorname{curl}_1 \mathbf{A} = \frac{1}{h_2 h_3} \left(\frac{\partial}{\partial q_2}\, A_3 h_3 - \frac{\partial}{\partial q_3}\, A_2 h_2 \right)$$

$$\operatorname{curl}_2 \mathbf{A} = \frac{1}{h_3 h_1} \left(\frac{\partial}{\partial q_3}\, A_1 h_1 - \frac{\partial}{\partial q_1}\, A_3 h_3 \right)$$

$$\operatorname{curl}_3 \mathbf{A} = \frac{1}{h_1 h_2} \left(\frac{\partial}{\partial q_1}\, A_2 h_2 - \frac{\partial}{\partial q_2}\, A_1 h_1 \right) \quad (11.47)$$

If it is necessary to form the Laplacian of a vector, the following procedure is adopted. Apply the operation of finding the gradient of the divergence of the vector according to (11.44) and (11.45). Then develop the curl of the curl by performing operation (11.47) twice. From (11.42), the difference between the two expressions yields the required Laplacian. Here, generally speaking, we find that some component of the Laplacian of a vector depends not only on the same component of the vector itself but on the other two as well.

Transformation to Tensor Notation. We shall now show how some vector operations are written in tensor notation. We always make use of the summation rule (see Part I, Section 9).

Then the divergence of a vector in orthogonal coordinates is written very simply:

$$\text{div}\,\mathbf{A} = \frac{\partial A_x}{\partial x} + \frac{\partial A_y}{\partial y} + \frac{\partial A_z}{\partial z} = \frac{\partial A_\alpha}{\partial x_\alpha} \tag{11.48}$$

Tensor operations in curvilinear coordinates are much more involved and will not be required in this book.

In tensor form the Laplacian of a scalar quantity is written as follows:

$$\nabla^2 \psi = \frac{\partial}{\partial x_\alpha} \frac{\partial}{\partial x_\alpha} \psi \equiv \frac{\partial^2 \psi}{\partial x_\alpha^2} \tag{11.49}$$

The Laplacian of a vector in Cartesian coordinates is written similarly.

We introduce the concept of an *invariant tensor*, as we call a tensor which retains its form in a transformation to another coordinate system. We have already encountered one such tensor: the tensor $\delta_{\mu\nu}$ in Section 9. Another tensor possessing this property is the tensor of rank 3, $\varepsilon_{\mu\nu\lambda}$, whose all components are equal to $+1$ if the indices are in cyclic order (123, 312, and 231), to -1 if the order is not cyclic (213, 132, and 321), and zero if any two of the indices μ, ν, λ are equal.

In other words, it can be said that the tensor $\varepsilon_{\mu\nu\lambda}$ is antisymmetric with respect to any pair of indices: a permutation of them reverses the sign of its component. Indeed, component *213* differs from *123* by one permutation and is therefore negative, while *312* is obtained from *123* by two permutations. Two sign reversals yield a plus. Obviously, permutations of the same indices affect nothing, but on the other hand, it should reverse the sign of the component. Hence a component with identical indices is equal to zero, since only zero is equal to itself with its sign reversed.

The property of antisymmetry of a tensor with respect to any pair of indices is conserved in a rotation of the coordinate system or,

in other words, it is an invariant property. Indeed, the equality $\varepsilon_{\mu\nu\lambda} = -\varepsilon_{\mu\nu\lambda}$ is of tensor form and is therefore valid in any coordinate system. This can be verified directly by applying the tensor transformation formulas.

But it also follows from this that the same components of tensor $\varepsilon'_{\mu\nu\lambda}$ differ from zero as those of tensor $\varepsilon_{\mu\nu\lambda}$ in the initial coordinates system, and that they are of the same sign. It remains to show that the components of $\varepsilon'_{\mu\nu\lambda}$ are also equal to ± 1.

We write the general transformation formula of a tensor:

$$\varepsilon'_{\mu\nu\lambda} = (\mu,\ \alpha)\,(\nu,\ \beta)\,(\lambda,\ \gamma)\,\varepsilon_{\alpha\beta\gamma} \tag{11.50}$$

But we have seen that the symmetry of the tensor $\varepsilon'_{\mu\nu\lambda}$ is the same as that of the initial tensor $\varepsilon_{\alpha\beta\gamma}$. Hence, $\varepsilon'_{\mu\nu\lambda} = C\varepsilon_{\mu\nu\lambda}$, where C is a number that has to be determined. We introduce it into (11.50) and multiply both sides of the equation by $\varepsilon_{\mu\nu\lambda}$. In the left-hand side we have C multiplied by the sum of the squares of the components, $\varepsilon_{\mu\nu\lambda}\varepsilon_{\mu\nu\lambda}$. There are altogether six such components, which we listed before. In the right-hand side we have a determinant made up of the transformation coefficients $(\mu,\ \alpha)$ which is also multiplied by 6.

Indeed, the coefficients $(\mu,\ \alpha)$, $(\nu,\ \beta)$, and $(\lambda,\ \gamma)$ are taken each time in threes and, according to the properties of $\varepsilon_{\alpha\beta\gamma}$, only from different rows and columns of the table, with the sign corresponding to whether the permutation of the rows and columns is even or odd. The factor 6 is obtained because in the summation one triplet of indices can be taken arbitrarily, while the number of permutations of three by three is 3!, or 6.

After cancelling out 6 we find that the required number C is equal to a determinant made up of the coefficients of the rotation, which is equal to unity. Hence $\varepsilon_{\mu\nu\lambda}$ is an invariant tensor which retains its form in any rotation of the coordinate system.

We now form the following combination from two vectors and a tensor:

$$D_\alpha = \varepsilon_{\alpha\beta\gamma}A_\beta B_\gamma \tag{11.51}$$

The quantity D_α written in terms of its components has the following values:

$$D_1 = A_2 B_3 - A_3 B_2,\ \ D_2 = A_3 B_1 - A_1 B_3,\ \ D_3 = A_1 B_2 - A_2 B_1$$

We have obtained the components of a vector product. Equation (11.51) immediately shows that a vector product behaves like a vector in a transformation of the coordinates, because (11.51) is a tensor equation: such equations are valid in any coordinate system and consequently lack the arbitrariness associated with its choice. Any equation expressing a physical law must possess this

property, which is the main reason for the use of vector and tensor equations in physics.

The curl of the vector is determined similarly to (11.51):

$$\varepsilon_{\alpha\beta\gamma}\frac{\partial A_\gamma}{\partial x_\beta} = (\nabla\times A)_\alpha = \text{curl}_\alpha\, A \qquad (11.52)$$

EXERCISES

1. Use Eqs. (11.26)-(11.42) to calculate the following expressions without introducing components: (a) $\nabla^2\,(1/r)$ at $r \neq 0$; (b) div $\varphi\,(r)\,\mathbf{r}$, curl $\varphi\,(r)\,\mathbf{r}$; (c) $\nabla\,(\mathbf{A}\cdot\mathbf{r})$, where $\mathbf{A}$ is a constant; (d) $\nabla\,(\mathbf{A}\,(r)\cdot\mathbf{r})$; (e) div $\varphi\,(r)\,\mathbf{A}\,(r)$, curl $\varphi\,(r)\,\mathbf{A}\,(r)$; (f) div $[\mathbf{r}\times(\mathbf{A}\times\mathbf{r})]$, where $\mathbf{A} = $ constant; (g) curl $\mathbf{r}[\times(\mathbf{A}\times\mathbf{r})]$, where $\mathbf{A} = $ constant; (h) $\nabla^2\mathbf{A}\,(r)$, see (11.42); (i) $\nabla\,(\mathbf{A}\,(r)\cdot\mathbf{B}\,(r))$; (j) curl $(\mathbf{A}\times\mathbf{r})$, where $\mathbf{A} = $ constant; (k) div $(\mathbf{A}\times\mathbf{r})$, where $\mathbf{A} = $ constant; (l) $\nabla^2\,(\mathbf{r}/r)$.

Answers. (a) $\nabla^2\,(1/r) = $ div grad $(1/r) = -$div $(\mathbf{r}/r^3) = 0$; (b) $3\varphi + r\dot\varphi$ and 0; (c) $\mathbf{A}$; (d) $\mathbf{A} + \mathbf{r}\,(\mathbf{r}\cdot\dot{\mathbf{A}})/r$; (e) $\dot\varphi\,(\mathbf{r}\cdot\mathbf{A})/r + \varphi\,(\mathbf{r}\cdot\dot{\mathbf{A}})/r$ and $\dot\varphi\,(\mathbf{r}\times\mathbf{A})/r + \varphi\,(\mathbf{r}\times\dot{\mathbf{A}})/r$; (f) $-2\,(\mathbf{A}\cdot\mathbf{r})$; (g) $3\,(\mathbf{r}\times\mathbf{A})$; (h) $\ddot{\mathbf{A}} + 2\dot{\mathbf{A}}/r$; (i) $\mathbf{r}\,(\mathbf{A}\cdot\dot{\mathbf{B}})/r + \mathbf{r}\,(\dot{\mathbf{A}}\cdot\mathbf{B})/r$; (j) $2\mathbf{A}$; (k) 0; (l) $-2\mathbf{r}/r$.

2. Write $\nabla^2\psi$ in cylindrical coordinates.

3. Write the three components of $\nabla^2\mathbf{A}$ in spherical coordinates.

4. Prove that $\varepsilon_{\alpha\beta\gamma}\varepsilon_{\alpha\mu\nu} = \delta_{\beta\mu}\delta_{\gamma\nu} - \delta_{\beta\nu}\delta_{\gamma\mu}$, and deduce from this the rule $\mathbf{A}\times(\mathbf{B}\times\mathbf{C}) = \mathbf{B}\,(\mathbf{A}\cdot\mathbf{C}) - \mathbf{C}\,(\mathbf{A}\cdot\mathbf{B})$.

5. Solve the tensor equation $x_\alpha x_\beta A_\beta - A_\alpha = a x_\alpha$, where x_α is the unknown vector, and vector A_β and scalar a are given.

Solution. Multiply both sides of the equation by A_α, which leads to a quadratic equation for the scalar $A_\alpha x_\alpha$. Solving it, we substitute the result into the initial equation, which yields a linear equation with respect to the components of vector x_α.

6. Given a straight line and a point O at a distance a from it. Let a point in the plane through the line and the given point O be at a distance z from the line and r from point O. Introducing the coordinates $\xi = (r + z)/2$, $\eta = (r - z)/2$, and φ, where φ is the angle of rotation around an axis through O perpendicular to the straight line, write the Laplacian in terms of these coordinates.

12

MAXWELL'S EQUATIONS

Interaction in Mechanics and in Electrodynamics. Interactions in electrodynamics take place not between individual charges, but between charges and the surrounding electromagnetic field. The physical concept of a field in electrodynamics differs essentially from the field concept in Newtonian mechanics.

We know that the space in which gravitational forces act is called a gravitational field. The value of these forces at any point of the field is determined, in Newtonian mechanics, by the instantaneous positions of the gravitating bodies, no matter how far they are from the given point. In electrodynamics such a field representation is not satisfactory: during the time that it takes an electromagnetic disturbance to move from one charge to another, the latter can move a very great distance. Elementary charges (electrons, protons, mesons) very often have velocities close to the velocity of propagation of electromagnetic disturbances.

The modern theory of gravity (the general theory of relativity) shows that gravitational interaction, too, propagates with a finite velocity. But since macroscopic bodies move considerably slower, within the scale of the solar system the finite velocity of propagation of gravitational forces introduces only an insignificant correction to the laws of motion of Newtonian mechanics.

In the electrodynamics of elementary charges, the finite velocity of propagation of electromagnetic disturbances is of fundamental significance. If the action of a field affects the energy or momentum of a charged particle, the change can be directly transmitted only to the surrounding electromagnetic field, because, for the energy and momentum of other particles to change, a finite time interval is required before the electromagnetic disturbance excited by the charge reaches them. But this means that the electromagnetic field itself possesses energy and momentum, otherwise these two important mechanical quantities would not always be conserved, vanishing at the instant when the signal is emitted and reappearing at the instant when it is received.

In Newtonian mechanics it is assumed that a disturbance is transmitted instantaneously, hence there is no need to ascribe momentum or energy to the field: as soon as one gravitating particle releases a certain momentum or energy another immediately acquires them.

Since, as has just been pointed out, an electromagnetic field possesses momentum and energy, it can be treated as an independent physical entity in exactly the same way as charged particles. The equations of electrodynamics must directly describe the propagation

of electromagnetic disturbances in space and the interactions of the charges with the field.

Interaction between charges is effected through the electromagnetic field. Such laws as the Coulomb or Biot-Savart laws, in which only instantaneous positions or instantaneous velocities of the charges appear, are of an approximate nature and hold only when the relative velocities of the charges are small compared with the propagation velocity of the electromagnetic disturbances. This velocity is a fundamental constant appearing in the equations of electrodynamics. It is to a great degree of accuracy equal to $3 \times \times 10^{10}$ cm-s^{-1}.

The reality of the field is particularly evident from the fact that electrodynamic equations admit of a solution in the absence of charges. These solutions describe electromagnetic waves in vacuum, in particular light and radio waves.

For two centuries, the supporters of the wave theory of light considered that light waves were propagated by a special elastic medium permeating all space, the so-called "ether". In order to represent the spread of oscillations it was, naturally, necessary to have something oscillating. This "something" was called the *ether*. Proceeding from an analogy with the propagation of sound waves in a continuous medium, the ether was endowed with the properties of a fluid, physical phenomena being explained simply by reducing them to definite mechanical displacements of bodies. In particular, light phenomena were regarded as displacements of particles of the special medium, the ether.

The enunciation of the electromagnetic theory of light led physicists to the conclusion that the electromagnetic field is real in the same sense as matter. Moreover, the laws of electrodynamics should form the basis for the deduction of the complex laws governing the motions of the atoms of matter, in particular, fluids. The bearer of electromagnetic field is physical space, which is inseparable from the states and motions of real entities.

Electromagnetic Field. The investigation of the electromagnetic field began with its most apparent manifestations: the electric force produced by the rubbing of bodies, the properties of magnets, and the like. It was known for very long that bodies are divided into conductors and insulators, that breaking a magnet in half yields not two separate poles but two magnets with two poles each, etc.

At the time, physicists were unable to explain why some bodies conduct electricity while others do not, why iron is strongly magnetic and copper is evidently not at all. Nevertheless, physicists were able, without taking up such problems, to learn some of the basic laws of electromagnetism, such, for example, as the Coulomb law of the interaction force between two charges, which does not depend

on the origin of the electrification of the bodies carrying the charges. The same can be said of Faraday's law of electromagnetic induction, which relates magnetic and electric forces in a form that does not reflect the properties of specific bodies.

In the first volume of this book we shall deal exclusively with the fundamental laws of electromagnetism. To formulate them there is no need to consider the properties of ponderable media, like metals, dielectrics, or ferromagnetics. We shall leave specific media alone for the time being and deal here only with separate charges and the fields they produce.

This form of presentation of electrodynamics was formerly known as the "electron theory", on the assumption that an electron is simply a point charge and nothing more. A real electron possesses very complex properties, and though it in no way manifests its actual dimensions in experiments, it bears little semblance to the idealized charge of the "electron theory".

In a sense the charge we shall be speaking of in presenting the subject-matter of this part of the book resembles the "mass point" of Newtonian mechanics. It is an idealized entity, for the present most suitable for the formulation of fundamental theoretical laws.

In developing Maxwell's equations—the fundamental equations of electrodynamics—we shall, for a time, have to make use of terms borrowed from the theory dealing with the properties of material media, such as the concept of a current-carrying circuit. Actually, though, we do not take account of the complex properties of real conductors. In fact, the term "conductor" is employed only for considerations of physical visualization: we have in mind an imaginary circuit along which a continuously distributed charge is moving.

One remark of a terminological nature is called for. We shall everywhere say simply "field" instead of "field strength". This is conventional in theoretical physics.

We shall also use the Gaussian system of units, in which electric and magnetic fields are expressed in quantities of the same dimensions, $g^{1/2}cm^{-1/2}s^{-1}$. [2]

Electromotive Force. We shall begin with the definition of electromotive force in a circuit: this is the work done by the forces of an electric field in the passage of a unit charge along a given circuit; it is absolutely immaterial whether the circuit is filled with a conductor or is simply a closed line drawn in space. In the latter sense

---

[2] This choice of units is especially convenient in the theory of relativity (see Sec. 15), in which electric and magnetic field components or the absolute values of the fields are involved in the form of linear combinations. The reader will find more about different system of units in the book *Units of Physical Quantities and Their Dimensions*, by L. A. Sena (Mir Publishers, Moscow, 1972).

the concept of electromotive force can be applied to a cyclic induction accelerator of electrons, the betatron.

Let us write the expression of electromotive force (abbreviated as emf and denoted $\mathscr{E}$) in the notation of Section 11. The force acting on a unit charge at a given point is, by definition, the electric field **E**. The work done by this force on an element of path dl is the scalar product **E** dl. Then the work done on the whole closed circuit, or the emf, is equal to the integral

$$\mathscr{E} = \int \mathbf{E}\, d\mathbf{l} \tag{12.1}$$

Suppose a surface is stretched over a given circuit. Denoting the magnetic field by **H**, we find that the magnetic flux across an element of the surface is, by the definition given in Section 11, $d\Phi = \mathbf{H}\, d\mathbf{S}$. The magnetic flux across the whole surface stretched on the circuit is

$$\Phi = \int \mathbf{H}\, d\mathbf{S} \tag{12.2}$$

It is important that the magnitude of the flux, Φ, does not depend on the specific form of the surface stretched on the circuit. This can be visually explained by the fact that the magnetic field lines cannot originate or terminate in an empty space devoid of magnets. Consequently, if two different surfaces are stretched over the circuit, the flux across each must be the same—it can neither decrease nor increase between them.

Faraday's induction law is written in the form of the following equation:

$$\mathscr{E} = -\frac{1}{c}\frac{\partial \Phi}{\partial t} \tag{12.3}$$

If all quantities are expressed in the Gaussian system, the proportionality factor is equal to 3×10^{10} cm-s^{-1}. It is readily apparent from (11.1) and (11.2) that c has the dimensions of velocity.

If a circuit is in vacuum, as in the case of an induction accelerator, the work done on the charge augments its energy.

Maxwell's Equation for curl E. Thus, Eq. (12.3) refers to any arbitrary closed circuit. We substitute Eqs. (12.1) and (12.2) into this equation to get

$$\int \mathbf{E}\, d\mathbf{l} = -\frac{1}{c}\frac{\partial}{\partial t}\int \mathbf{H}\, d\mathbf{S} \tag{12.4}$$

The left-hand side of the equation can be transformed by Stokes' theorem (11.19), and in the right-hand side the order of time differentiation and surface integration can be interchanged, since they are performed for independent variables. In addition, taking this

integral over to the left-hand side, we obtain

$$\int \left(\operatorname{curl} \mathbf{E} + \frac{1}{c} \frac{\partial \mathbf{H}}{\partial t} \right) d\mathbf{S} = 0 \tag{12.5}$$

But, the initial circuit is completely arbitrary, that is, it can have arbitrary magnitude and shape. Let us assume that the expression in parentheses in (12.5) is not equal to zero. Then we can choose the surface and the circuit that bounds it so that the integral (12.5) does not become zero. Thus, the following equation must be satisfied:

$$\operatorname{curl} \mathbf{E} + \frac{1}{c} \frac{\partial \mathbf{H}}{\partial t} = 0 \tag{12.6}$$

In comparison with (12.3), this equation does not contain anything new physically; it is the same induction law, but rewritten in differential form for an infinitely small circuit. In many applications the differential form is more convenient than the integral form.

The Equation for div H. As we have already said, magnetic lines of force cannot originate or terminate in vacuum, that is, they are either closed or go off to infinity. Hence, into any closed surface the same number of magnetic field lines enter as leave. The magnetic flux in free space, across any closed surface, is equal to zero:

$$\int \mathbf{H} \, d\mathbf{S} = 0 \tag{12.7}$$

Transforming this integral to a volume integral according to the Gauss theorem (11.6), we obtain

$$\int \operatorname{div} \mathbf{H} \, dV = 0 \tag{12.8}$$

Since the surface bounding the volume is completely arbitrary, we can always choose this volume to be so small that the integral is taken over the region in which div **H** is of constant sign, if it is not equal to zero. But then, contrary to (12.7) and (12.8), div **H** will not be equal to zero. Therefore, the divergence of **H** must everywhere vanish:

$$\operatorname{div} \mathbf{H} = 0 \tag{12.9}$$

The expression (12.9) is the differential form of (12.7) for an infinitely small volume.

In Section 11 it was shown that the divergence of a vector is the density of sources of a vector field. The sources of the field may be free charges, as in the case of an electric field. But a magnetic field does not correspond to any free charges.

Equations (12.6) and (12.9) are together called the *first pair of Maxwell's equations.* Let us now introduce the second pair.

The Equation for div E. The electric flux across a closed surface is equal to the total electric charge inside the surface multiplied by 4π (the *Gauss law*)

$$\int \mathbf{E} \, d\mathbf{S} = 4\pi e \qquad (12.10)$$

This law is derived from the Coulomb law for point charges. The field due to a point charge e is expressed by the following equation:

$$\mathbf{E} = \frac{e}{r^2} \frac{\mathbf{r}}{r}$$

Here, $\mathbf{r}$ is a radius vector drawn from the point where the charge is located to the point where the field is defined. The field is inversely proportional to r^2 and is directed along the radius vector.

Let us surround the charge by a spherical surface centred on the charge. The element of surface for the sphere, $d\mathbf{S}$, is $r^2 \, d\Omega \mathbf{r}/r$, where $d\Omega$ is a solid-angle element, and $\mathbf{r}/r$ indicates the direction of the normal to the surface. The flux of the field across the surface element is

$$\mathbf{E} \, d\mathbf{S} = \frac{e}{r^2} \times \frac{\mathbf{r}}{r} \times r^2 \, d\Omega \, \frac{\mathbf{r}}{r} = e \, d\Omega$$

The flux across the whole surface of the sphere is $\int e \, d\Omega = e \int d\Omega = 4\pi e$. But since lines of force begin only at a charge, the flux will be the same through the sphere as through any closed surface around the charge. Therefore, if there is an arbitrary charge distribution e inside a closed surface, Eq. (12.10) holds.

In order to rewrite this equation in differential form, we introduce the concept of charge density. The *charge density* ρ is the charge contained in unit volume, so that the total charge in a volume is related to the density by the following equation:

$$e = \int \rho \, dV \qquad (12.11)$$

Hence, $\rho = \lim\limits_{\Delta V \to 0} \dfrac{\Delta e}{\Delta V}$. Introducing the charge density into (12.10),

$$\int (\text{div } \mathbf{E} - 4\pi\rho) \, dV = 0 \qquad (12.12)$$

Repeating the same reasoning for this integral as we applied to (12.8), we obtain an equation of similar form for an electric field:

$$\text{div } \mathbf{E} = 4\pi\rho \qquad (12.13)$$

According to (11.8), we can say that the density of sources of an electric field is equal to the electric charge density multiplied by 4π.

In formulating the fundamental laws of electrodynamics it is often convenient to treat charges as point charges (remembering that these charges should not be physically identified with electrons!). For point charges the density function is given by means of a limiting process.

Let us first assume that a charge of finite magnitude is distributed over a small, but also finite, volume ΔV. Then ρ must be regarded as the ratio $e/\Delta V$. If we let the volume ΔV tend to zero, then the density function will have a very peculiar form: it will turn out to be equal to zero everywhere except at the place where the charge is situated, and at that point it will convert to infinity, since the numerator of the fraction $e/\Delta V$ is finite and the denominator is infinitely small. However, the integral

$$\int \rho\, dV = \frac{e}{\Delta V} \int dV = \frac{e}{\Delta V}\, \Delta V = e$$

remains equal to the charge e.

Thus, the concept of charge density can also be used in the case of a point charge. In this case ρ is understood to be a function which is equal to zero everywhere except at the point of the charge. The volume integral of this function is either equal to the charge e itself, if the charge is situated inside the integration region, or zero, if the charge is outside the region of integration.

The Law of Charge Conservation. One of the most important laws of electrodynamics is the law of conservation of charge: the total charge of any system remains constant if no external charges are brought into it. In all charge transformations occurring in nature, the law of conservation of charge is satisfied with extreme precision (while the law of conservation of mass is approximate).

In order to formulate the charge-conservation law in differential form, we must introduce the concept of *current density*. This vector quantity is defined as

$$\mathbf{j} = \rho\mathbf{v} \tag{12.14}$$

where $\mathbf{v}$ is the charge velocity at the point where the density ρ is defined. The dimension of charge density is charge/cm³, and of current density, charge/cm²s (that is, the dimensions of charge passing in unit time across unit area). In particular, for the point charge in Eq. (12.14), $\mathbf{v}$ denotes its velocity, and ρ the density function defined above.

The total current emerging from an area is

$$I = \int \mathbf{j}\, d\mathbf{S} \tag{12.15}$$

According to the charge conservation law, I must be equal to the reduction of charge inside the surface in unit time:

$$I = -\frac{\partial e}{\partial t} \qquad (12.16)$$

As we have already done in other cases, we pass from this integral form of the charge-conservation law to the differential form. Substituting e from (12.11) and transforming I by the Gauss theorem, we obtain

$$\int \left(\frac{\partial \rho}{\partial t} + \operatorname{div} \rho \mathbf{v} \right) dV = 0 \qquad (12.17)$$

Since the volume over which the integration is performed is arbitrary, the charge-conservation law in differential form follows from (12.17):

$$\frac{\partial \rho}{\partial t} + \operatorname{div} \mathbf{j} = \frac{\partial \rho}{\partial t} + \operatorname{div} \rho \mathbf{v} = 0 \qquad (12.18)$$

Displacement Current. From direct-current theory it is known that current lines are always closed. Indeed, open lines indicate that there is either an accumulation or loss of charge at their ends. But we can also define vector lines such that they will always be closed or go to infinity in the case of alternating currents. For this we substitute the derivative $\partial \rho/\partial t$ according to (12.13) into the equation of the charge-conservation law (12.18). This derivative is equal to $(1/4\pi) \operatorname{div} (\partial \mathbf{E}/\partial t)$. Hence, we always have the relation

$$\operatorname{div} \left(\mathbf{j} + \frac{1}{4\pi} \frac{\partial \mathbf{E}}{\partial t} \right) = 0 \qquad (12.19)$$

Comparing (12.19) and (12.9), we see that the vector lines

$$\mathbf{j} + \frac{1}{4\pi} \frac{\partial \mathbf{E}}{\partial t}$$

are always closed. The vector $(1/4\pi)(\partial \mathbf{E}/\partial t)$ is called the *displacement current*, because it is not associated with the transfer of charge.[3] Together with the charge-transport current, the displacement current forms a closed system of vector lines.

So far all we have done is to rewrite in a somewhat different form the laws of electrodynamics known from elementary physics. Now we must introduce a substantially new assumption: the magnetic action of the displacement current differs in no way from the magnetic action of the charge-transfer current. This was Maxwell's assumption when he formulated the general laws of electrodynamics.

The Equation for curl H. Equation (12.13) belongs to the second pair of Maxwell's equations. Another equation of this pair defines curl **H**. Just as in the development of (12.13) we made use of the

---

[3] In future we shall use simply "current" instead of "current density".

Coulomb law, here we shall require the law of Biot-Savart and the hypothesis of the magnetic action of the displacement current.

First we write the Biot-Savart law in elementary form. Let a current element of length dl_1 (Figure 20) be located at a point whose radius vector is r_1, the current strength being I. Then at a point with radius vector r this current produces a field dH defined by the law

$$dH = \frac{I}{c} \frac{dl_1 \times (r - r_1)}{|r - r_1|^3} \tag{12.20}$$

It is more convenient to deal with a spatially distributed rather than a linear current. If the current density is ρv (neglecting, for
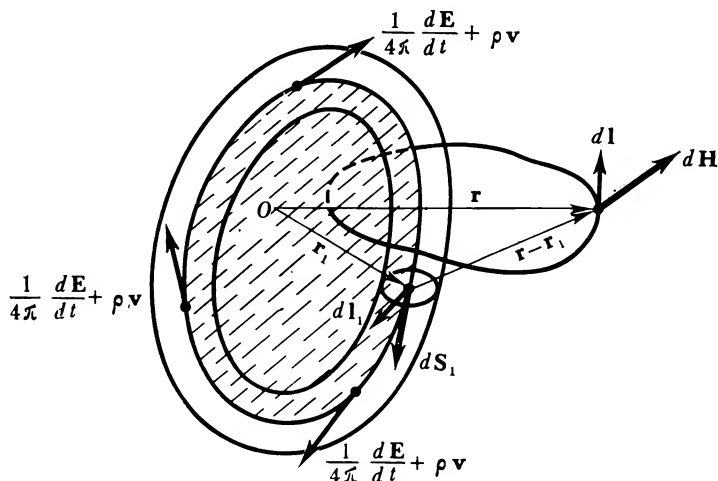
Figure 20

the time being, the displacement current), the intensity can be written as the total flux across an area dS:

$$I = \rho \, (v \cdot dS) \tag{12.21}$$

Hence, to obtain the magnetic field of the whole distribution of currents, the fundamental law (12.20) must be integrated over a certain area crossed by the current and along the whole circuit dl_1:

$$H = \int \rho \, (v \cdot dS) \int \frac{dl_1 \times (r - r_1)}{c \, |r - r_1|^3} \tag{12.22}$$

This yields the portion of the magnetic field due to the currents passing along the outer circuit enclosing this area (see Figure 20). If we add the displacement current $(1/4\pi)(\partial E/\partial t)$ to ρv, the vector lines of the resultant current will close. In Figure 20 these closed

lines cross the unhatched circuit. The points of this circuit and the surface stretched across it are defined by the radius vector $\mathbf{r}$, while the points of the vector line of the total current and the corresponding hatched surface are defined by the radius vector $\mathbf{r_1}$. Note that if we hadn't introduced the displacement current, the paths of point charges could at no given instant be closed: the current at that instant is not zero where there is a charge but it is zero at all other points in space. Fully closed lines, as in Figure 20, are obtained only together with the displacement current.

Now let us integrate the expression for the magnetic field along the contour of the unhatched surface, that is, over $d\mathbf{l}$. Then, from (12.22), after adding $(1/4\pi)\,(\partial\mathbf{E}/\partial t)$ and rearranging the order of integration, we obtain

$$\int \mathbf{H}\,d\mathbf{l} = \frac{1}{c}\int \left(\rho\mathbf{v} + \frac{1}{4\pi}\frac{\partial\mathbf{E}}{\partial t}\right) dS \int d\mathbf{l} \int \frac{d\mathbf{l_1}\times(\mathbf{r}-\mathbf{r_1})}{|\mathbf{r}-\mathbf{r_1}|^3} \quad (12.23)$$

Let us now show that the inner double integral

$$A = \int d\mathbf{l} \int \frac{d\mathbf{l_1}\times(\mathbf{r}-\mathbf{r_1})}{|\mathbf{r}-\mathbf{r_1}|^3}$$

is equal to 4π, if the hatched contour and the vector line contour along which integration over $d\mathbf{l_1}$ is performed are connected; otherwise it is equal to zero.

We denote differentiation with respect to the components of vector $\mathbf{r_1}$ by the symbol ∇_1, and differentiation with respect to the components of $\mathbf{r}$ by the symbol ∇. Since the integrand depends only upon the difference $\mathbf{r}-\mathbf{r_1}$, we can write symbolically $\nabla_1 = -\nabla$. We also replace $(\mathbf{r}-\mathbf{r_1})/|\mathbf{r}-\mathbf{r_1}|^3$ by $\nabla_1|\mathbf{r}-\mathbf{r_1}|^{-1}$:

$$A = \int d\mathbf{l} \int \left(d\mathbf{l_1}\times\nabla_1\frac{1}{|\mathbf{r}-\mathbf{r_1}|} \right) \quad (12.24)$$

We perform a cyclic permutation in the mixed product so that the element of length $d\mathbf{l_1}$ would appear *after* the sign of the vector product:

$$A = \int\int d\mathbf{l_1} \left(\nabla_1\frac{1}{|\mathbf{r}-\mathbf{r_1}|}\times d\mathbf{l} \right) \quad (12.25)$$

We have obtained an integral along the circuit $d\mathbf{l_1}$ to which Stokes' theorem can be applied, so that

$$A = \int\int \mathrm{curl}_1\left(\nabla_1\frac{1}{|\mathbf{r}-\mathbf{r_1}|}\times d\mathbf{l} \right) dS_1 \quad (12.26)$$

Here the integration is taken over the surface bounded by the current line. Using (11.30) and remembering that $d\mathbf{l}$ is not differentiated, we expand the curl of the vector product. As a result

$\operatorname{div} \nabla_1 \mid \mathbf{r} - \mathbf{r}_1 \mid^{-1} = \nabla_1^2 \mid \mathbf{r} - \mathbf{r}_1 \mid^{-1}$ vanishes, leaving only

$$- \int d\mathbf{l} \, \nabla \int \frac{(\mathbf{r} - \mathbf{r}_1) \, d\mathbf{S}_1}{\mid \mathbf{r} - \mathbf{r}_1 \mid^3} \tag{12.27}$$

The quantity under the inner integral sign has a direct geometrical meaning. Draw a straight line from a point on the circuit $d\mathbf{l}$ to its intersection with surface S_1. The vector $\mathbf{r} - \mathbf{r}_1$ lies along this line. The product of its multiplication by $d\mathbf{S}_1$ divided by $\mid \mathbf{r} - \mathbf{r}_1 \mid$ denotes the projection of the area $d\mathbf{S}_1$ on a surface perpendicular to $\mathbf{r} - \mathbf{r}_1$. Hence, the expression

$$\frac{1}{\mid \mathbf{r} - \mathbf{r}_1 \mid^2} \frac{(\mathbf{r} - \mathbf{r}_1) \, d\mathbf{S}_1}{\mid \mathbf{r} - \mathbf{r}_1 \mid}$$

is a solid-angle element at which the area $d\mathbf{S}_1$ is seen from the point on the circuit $d\mathbf{l}$. The inner integral in (12.27) thus represents the total solid angle Ω at which the path described by the current line is seen from the point on the circuit $d\mathbf{l}$.

In writing (12.27) we replaced ∇_1 by ∇ and took it outside the integral over $d\mathbf{S}_1$. The product $d\mathbf{l} \, \nabla$ is the differential $d\Omega$ taken along the circuit l. Now let us calculate the integral along the circuit l. Suppose the current line was traced clockwise. Then the positive side of the surface S_1 stretched over this circuit lies under the page of Figure 20. If we look at the circuit of S_1, as it is shown in the drawing, the scalar product $(\mathbf{r} - \mathbf{r}_1) \, d\mathbf{S}_1$ is positive, and the solid angle must be taken with the plus sign.

Let the initial point of traversing circuit l be at the intersection of the circuit with surface S_1. We start the traverse on the side seen in the drawing. At this point the surface occupies the half-space seen from it, and therefore $\Omega = 2\pi$. Moving along the circuit l, we gradually reduce the solid angle, so that $d\Omega < 0$. On the reverse side of the surface S_1 the scalar product is negative, and the solid angle is equal to -2π. The total variation of the solid angle is $2\pi - (-2\pi)$, or 4π, taking into account the minus sign in (12.27).

This is the case of a circuit connected with the current line; for a circuit that is not connected with it the reversion is to the same point, with $\Omega = 2\pi$, and the integral becomes zero. It follows that the integral (12.23) involves the contribution of only those current lines that are connected with the circuit l, which is therefore equal to

$$\int \mathbf{H} \, dl = \frac{4\pi}{c} \int \left(\rho \mathbf{v} + \frac{1}{4\pi} \frac{\partial \mathbf{E}}{\partial t} \right) d\mathbf{S} \tag{12.28}$$

Now, transforming the left-hand side of (12.28) according to Stokes' theorem, and taking into account that the surface over which the integration is carried out is arbitrary, we arrive at Maxwell's equation for curl $\mathbf{H}$:

$$\operatorname{curl} \mathbf{H} = \frac{4\pi}{c} \, \mathbf{j} + \frac{1}{c} \frac{\partial \mathbf{E}}{\partial t} \tag{12.29}$$

It will be readily observed that this equation is in agreement with the law of charge conservation. Indeed, let us perform the divergence operation on it. From (11.40), div curl $\mathbf{H} = 0$, so that only the divergence of the right-hand side remains. If we substitute div $\mathbf{E}$ according to (12.13), we again arrive at (12.18), that is, to the law of charge conservation:

$$\frac{4\pi}{c}\,\text{div}\,\mathbf{j} + \frac{1}{c}\,\frac{\partial}{\partial t}\,\text{div}\,\mathbf{E} = \frac{4\pi}{c}\left(\text{div}\,\mathbf{j} + \frac{\partial\rho}{\partial t}\right) = 0$$

Equation (12.29) is not just a way of writing the Biot-Savart law in differential form. It involves the displacement current, which does not appear in direct-current theory.

Maxwell's Equations. Let us once again write the set of Maxwell's equations in the form derived from the fundamental laws of electromagnetism.

The first pair:

$$\text{curl}\,\mathbf{E} = -\frac{1}{c}\,\frac{\partial\mathbf{H}}{\partial t} \tag{12.30}$$

$$\text{div}\,\mathbf{H} = 0 \tag{12.31}$$

The second pair:

$$\text{curl}\,\mathbf{H} = \frac{1}{c}\,\frac{\partial\mathbf{E}}{\partial t} + \frac{4\pi\mathbf{j}}{c} \tag{12.32}$$

$$\text{div}\,\mathbf{E} = 4\pi\rho \tag{12.33}$$

In these equations we consider ρ and $\mathbf{j}$, that is, the time dependent charge and current distributions in space, to be known. The unknowns to be determined are both electromagnetic field components, $\mathbf{E}$ and $\mathbf{H}$. Each of them has three vector components.

In spite of the fact that both pairs form, together, eight equations, only six of them are independent, according to the number of field components. Indeed, the three components of each curl are constrained by div curl $= 0$, and, hence, are not independent of one another.

It was shown in the first part of this book that the laws of mechanics can be derived from certain symmetry laws and the principle of least action. This approach reveals more vividly the totality of experimental facts underlying Newtonian mechanics. We can approach electrodynamics in a similar way. Then its fundamental laws will be found to be simple corollaries of very general regularities. The very procedure of developing Maxwell's equations is thereby greatly simplified.

Electromagnetic Potentials. We can introduce new unknown quantities into the equations of electrodynamics so that each equation

contains only one unknown. In this way the overall number of equations is reduced. These new quantities are called *electromagnetic potentials*.

We choose the potentials so that the first pair of Maxwell's equations is identically satisfied. In order to satisfy Eq. (12.31), it is sufficient to put

$$\mathbf{H} = \text{curl } \mathbf{A} \tag{12.34}$$

where **A** is a vector called the *vector potential*. Then, according to (11.40), div **H** will be equal to zero identically. The electric field should be represented in the form

$$\mathbf{E} = -\frac{1}{c}\frac{\partial \mathbf{A}}{\partial t} - \text{grad } \varphi \tag{12.35}$$

where φ is a quantity called the *scalar potential*. From (11.39), curl grad $\varphi = 0$. Substitution of Eqs. (12.34) and (12.35) into (12.30) leads to an identical cancelling out.

The electromagnetic field vectors are the physically determinate quantities, insofar as they are involved in the expressions of forces acting on the charges and currents. The field strengths, or simply fields, as is conventionally said, are obtained from the potentials by differentiation. Therefore, the potentials are determined up to expressions that cancel out in the differentiation. It is natural to select the expressions in such a way as to make the potential equations as simple as possible. Let us first find the most general transformation of potentials which does not change the fields in Eqs. (12.34) and (12.35).

From Eq. (12.34) it can be seen that if we add the gradient of any arbitrary function to the vector potential, the magnetic field does not change, since the curl of a gradient is identically zero. Putting

$$\mathbf{A} = \mathbf{A}' + \text{grad } f(\mathbf{r}, t) \tag{12.36}$$

we see that the magnetic field, expressed in terms of such a modified potential, remains unchanged:

$$\mathbf{H} = \text{curl } \mathbf{A} = \text{curl } \mathbf{A}'$$

In order that the addition of grad f to the vector potential should not affect the electric field, we must also change the scalar potential:

$$\varphi = \varphi' - \frac{1}{c}\frac{\partial f}{\partial t} \tag{12.37}$$

where f is the same function as in (12.36). Then for the electric field we obtain

$$\mathbf{E} = -\frac{1}{c}\frac{\partial \mathbf{A}}{\partial t} - \operatorname{grad} \varphi$$

$$= -\frac{1}{c}\frac{\partial \mathbf{A'}}{\partial t} - \frac{1}{c}\frac{\partial}{\partial t}\operatorname{grad} f - \operatorname{grad} \varphi' + \operatorname{grad}\frac{1}{c}\frac{\partial f}{\partial t}$$

$$= -\frac{1}{c}\frac{\partial \mathbf{A'}}{\partial t} - \operatorname{grad} \varphi'$$

Consequently, the electric field does not change either. Thus, the potentials are determined to the accuracy of the transformations (12.36) and (12.37), which are called the *gauge transformations.*

The Equations for Potentials. Let us now choose an arbitrary function such that the second pair of Maxwell's equations leads to equations for the potentials of the simplest possible form. Substitution of (12.34) and (12.35) into (12.32) gives

$$\operatorname{curl} \operatorname{curl} \mathbf{A} = -\frac{1}{c^2}\frac{\partial^2 \mathbf{A}}{\partial t^2} - \frac{1}{c}\frac{\partial}{\partial t}\operatorname{grad} \varphi + \frac{4\pi \mathbf{j}}{c} \qquad (12.38)$$

We express curl curl A with the aid of (11.42). Then (12.38) is reduced to the following form:

$$-\nabla^2 \mathbf{A} + \frac{1}{c^2}\frac{\partial^2 \mathbf{A}}{\partial t^2} + \operatorname{grad}\left(\operatorname{div} \mathbf{A} + \frac{1}{c}\frac{\partial \varphi}{\partial t}\right) = \frac{4\pi \mathbf{j}}{c} \qquad (12.39)$$

We shall now try to eliminate the quantity inside the parentheses in the left-hand side of (12.39). We denote it for brevity by a, and we perform the transformations (12.36) and (12.37) on the potentials. Then the quantity a is reduced to the form

$$a = \operatorname{div} \mathbf{A} + \frac{1}{c}\frac{\partial \varphi}{\partial t} = \operatorname{div} \mathbf{A'} + \frac{1}{c}\frac{\partial \varphi'}{\partial t} + \nabla^2 f - \frac{1}{c^2}\frac{\partial^2 f}{\partial t^2} \quad (12.40)$$

The function f has, so far, remained arbitrary. Let us now assume that it has been chosen so as to satisfy the equation

$$\nabla^2 f - \frac{1}{c^2}\frac{\partial^2 f}{\partial t^2} = 0 \qquad (12.41)$$

Then, from (12.40), it is obvious that the potentials will be subject to the condition

$$\operatorname{div} \mathbf{A'} + \frac{1}{c}\frac{\partial \varphi'}{\partial t} = 0 \qquad (12.42)$$

This is called the *Lorentz condition.*

As was shown, the expression of fields in terms of potentials is not changed by gauge transformations. For this reason we shall in future always consider that these transformations are performed

so that the Lorentz condition is satisfied; the primes in the poten-
tials can then be omitted.

From the Lorentz condition and (12.39) we obtain the equation
for the vector potential:

$$\nabla^2 \mathbf{A} - \frac{1}{c^2} \frac{\partial^2 \mathbf{A}}{\partial t^2} = -\frac{4\pi \mathbf{j}}{c} \qquad (12.43)$$

It is now also easy to obtain the equation for the scalar potential.
From (12.33) and (12.35) we have

$$\operatorname{div} \mathbf{E} = -\frac{1}{c} \frac{\partial}{\partial t} \operatorname{div} \mathbf{A} - \nabla^2 \varphi = 4\pi \rho$$

Substituting div **A** from the Lorentz condition (12.42), we obtain

$$\nabla^2 \varphi - \frac{1}{c^2} \frac{\partial^2 \varphi}{\partial t^2} = -4\pi \rho \qquad (12.44)$$

Equations (12.43) and (12.44) each contains one unknown. There-
fore, none of the equations for potential depend on the others, and
they can be solved separately. This is valid, however, only in Carte-
sian coordinates; in curvilinear coordinates different components
of **A** are involved in the same equations.

The equations for potentials are second-order equations with res-
pect to the coordinate and time derivatives. To solve such equations
the initial values must be stated not only of the potentials but of
their time derivatives as well.

As we shall see later, in very many cases it is necessary to use
equations involving not the electromagnetic fields but the poten-
tials that define them. But since potentials are not single-valued
and may, for the same electromagnetic fields, receive different
supplementary terms according to Eqs. (12.36) and (12.37), care
must be taken that the form of any equation involving potentials
does not change in gauge transformations. The thing is that these
transformations involve a completely arbitrary function f, which
can be chosen in any form. Obviously, this requirement does not
refer to the relations defining the potentials, that is (12.43) and
(12.44), the validity of which is connected with the choice of poten-
tials satisfying the Lorentz condition.

No physical result can be dependent on the choice of an arbitrary
function f on which no preliminary restrictions were imposed. The
Lorentz condition was selected only for the purpose of simplifying
(12.43) and (12.44); it is not physically necessary.

It is said that physical equations, that is, equations for directly
observable quantities, must be *gauge invariant*.

EXERCISE,

Write Maxwell's equations and the equations for potentials in cylindrical coordinates, for which $dl^2 = dr^2 + r^2\,d\psi^2 + dz^2$.

13

EINSTEIN'S RELATIVITY PRINCIPLE

Addition of Velocities and Electrodynamics. Maxwell's equations involve the constant c, which has the dimensions of velocity. It will be shown in Section 18 that c not merely happens to have the same dimensions as velocity but is in fact the speed with which electromagnetic waves propagate in vacuum. This result is obtained from a solution of the set (12.30)-(12.33), given the assumption that $\rho = 0$, $\mathbf{j} = 0$.

Suppose the quantities involved in these equations have been measured by an observer subject to inertial motion. Suppose, furthermore, that there is another observer moving relative to the first at a constant speed V. In mechanics the principle of relativity states that any equation must have the same form relative to different inertial frames of reference. In what conditions can this be valid in electrodynamics? Can the laws of electrodynamics be of the same form in the frame of reference of an initial observer, in which electromagnetic waves propagate in all directions with the speed c, and in the frame of another observer moving with a constant velocity V with respect to the first frame?

At first this seems impossible. According to the law of addition of velocities, in the second observer's frame of reference electromagnetic waves should propagate with a velocity

$$c' = c + V \tag{13.1}$$

where we have assumed, for the sake of simplicity, the waves to be propagating in the direction of the relative velocity of the reference frames. Reversal of the motion would reverse the sign in the equation; for the case of perpendicular motion the velocities would have to be added as a vector sum.

Thus, in the second observer's reference frame, which is also moving inertially, the propagation speed of electromagnetic waves would have to depend on direction. But since a travelling electromagnetic wave is one of the possible solutions of Maxwell's equa-

tions, they would have to be of different form in different inertial frames of reference.

It follows that only one of two statements can be true: either the conventional law of addition of velocities (13.1) is valid, and the relativity principle does not apply to electromagnetic fields, or the principle is valid for electromagnetism, and addition of velocities does not hold in its simplest form (13.1).

The Experiment of Michelson. The validity of the second assumption, that is, that the speed of light and, in general, of any electromagnetic disturbance in vacuum does not combine with the velocity
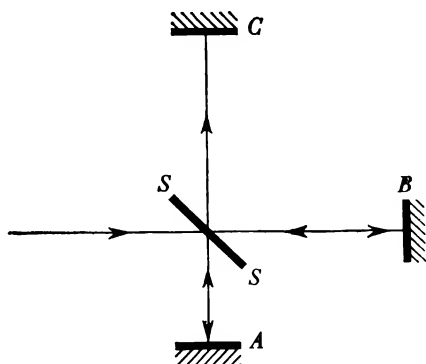
Figure 21

of a reference system, was demonstrated experimentally by Michelson in 1887. He showed that the velocity of light as measured in any inertial frame of reference is the same and equal to the fundamental constant c. Here is a brief description of the experiment.

A beam of light falls on a half-silvered mirror SS (Figure 21), where it is split in two: one part is reflected and falls on mirror A, the other passes through and falls on mirror B. Let beam SA be perpendicular to the velocity of the earth in its motion around the sun; then beam SB is parallel to that velocity. The light reflected from mirrors A and B returns to mirror SS; beam BS is reflected and falls on screen C, while beam AS passes through SS and falls on C directly. Both rays are thus entirely equivalent as regards their passage through, and reflections from, SS, but on sections AS and BS the light propagates differently relative to the earth's motion.

Let us see what effect could be expected if the velocity of light was compounded with the velocity of the earth according to the conventional law. Along the path SB the velocity of light relative to·

the earth would be $c - V$, and in the reverse direction $c + V$, where V is the velocity of the earth. In the assumption made, the time it takes for the light to travel the whole path SBS in both directions is

$$\frac{l}{c+V} + \frac{l}{c-V} = \frac{2lc}{c^2 - V^2} \approx \frac{2l}{c} + \frac{2lV^2}{c^3}$$

where $l = SB$. In going over to the approximate equation we made use of the fact that $V \ll c$.

Along section SA the velocities of the earth and light are perpendicular (in the reference frame fixed with respect to the apparatus). Assuming again that the velocity of light is added to that of the earth, we must now apply the vector law of addition. Then along section SA the velocity of light relative to the apparatus is equal to $(c^2 - V^2)^{1/2}$, c being the hypotenuse of the right triangle, and V and $(c^2 - V^2)^{1/2}$, its sides. The time it takes light to travel the whole path SAS, equal to $2l$, is

$$\frac{2l}{(c^2 - V^2)^{1/2}} \approx \frac{2l}{c} + \frac{lV^2}{c^3}$$

Thus, the difference between the times it takes light to travel along the paths SBS and SAS is equal to lV^2/c^3. By means of multiple reflections the paths travelled by the beams are made fairly long (tens of metres). By judicious manipulation of the paths the expected difference between the times it takes light to travel along SAS and SBS can be made equal to the half-period of the oscillation of light. Then, if our reasoning was correct, the beams on the screen should cancel out.

In order to make sure that the cancellation of the beams at a given point of the screen is due to the addition of the velocities of the earth and the light beams and not to some other causes, it is sufficient to rotate the apparatus through 45° so as to direct the velocity of the earth along the bisector of angle ASB. In this case the difference in time between the passage of the beams along SAS and SBS should in any case be zero, provided the difference in the initial position amounted to one-half the oscillation period. In other words, the interference fringes on the screen should move by half the distance between them, the light and dark areas interchange on the screen.

Actually, no change in the path difference of the beams in a rotation of the apparatus is observed, that is, the expected effect does not occur. The velocity of light is not added to the velocity of the earth.

Below we shall examine certain facts which would appear to point to a reverse conclusion and demonstrate that the contradiction is

an imaginary one. First we shall establish the corollaries that arise from the fact that addition of the velocity of light to the velocity of the earth does not occur.

The Relativity Principle Applied to Electromagnetic Field. As pointed out in the preceding section, the equations of electrodynamics do not presume the existence of an elastic medium—the "ether"—for the transmission of electromagnetic disturbances. The electromagnetic field is itself a reality, and it can therefore be expected that the equations of electromagnetics are just as independent of the choice of an inertial reference frame as the equations of mechanics. Both sets of equations directly describe motion, that is, the time rate of change of state of the observed entity. The form of the equations should not change depending on the chosen inertial frame of reference.

That is why the result of Michelson's experiment, far from contradicting the notion of the relativity of motion, confirms it. Michelson's experiment proves that the speed of light in vacuum is the same in all inertial frames of reference.

The velocity of propagation of interactions is a fundamental constant in the equations of electrodynamics. These equations can be invariant under transformations from one inertial system to another only when the velocity of propagation of interactions is the same in both systems. The result of Michelson's experiment contradicts only the law of addition of velocities, that is, the Galilean transformation (8.1).

This transformation law, and the consequent law of addition of velocities, is experimentally confirmed only in the case of relative velocities and velocities of motion that are small in comparison with the speed of light c. It follows from Michelson's experiment that for reference frames moving with high relative velocities and entities travelling at high speeds the Galilean transformation must be replaced by other, more exact transformations. Moreover, these transformations must be universal in form and the same for both particles and electromagnetic fields.

Indeed, let the charges in some specified inertial frame of reference interact in some way with an electromagnetic field, leading to certain events, for example, collisions of charges. Such events can be predicted on the basis of the equations of mechanics and electrodynamics. In a transformation to another inertial frame the equations of mechanics and electrodynamics must retain their form, as otherwise other effects would be observed, notably it could be found that the collision does not take place altogether. But collisions are objective facts which must be observable in all reference frames. Yet if we retain the Galilean transformations, even only for mechanics, and declare that the law of addition of velocities

consequent on them does not apply to electromagnetic fields, a discrepancy arises which is the greater the closer the velocity of the inertial system (or the velocities of motion of individual particles) approaches the speed of light.

We must replace the Galilean transformation by such transformations that would leave the equations of both mechanics and electrodynamics the same. However, in the process we find that the Newtonian laws of mechanics, valid only for small velocities of particles, have to be made more precise.

This calls for a revision of the laws of mechanics, however great our faith in them, based as it is on daily experience, with bodies whose speeds are small in comparison with the speed of light.

The Lorentz Transformation. We shall look for transformations of a more general form than the Galilean transformation for passing from one inertial frame of reference to another. Like the Galilean transformation, they must satisfy certain requirements of a general nature, listed below.

(i) The transformation equations must be symmetrical with respect to both inertial frames. We shall denote the quantities referring to one frame by unprimed letters, and those referring to the other, by primed letters. Thus, we must obtain equations expressing x, y, z, and t in terms of x', y', z', and t' in such a way that the primed quantities are expressed in terms of the unprimed ones, or the unprimed are expressed in terms of the primed ones, by equations of the same form.

Denoting the velocity of the primed frame relative to the unprimed frame V, we find that the direct transformation equations must then transform into the reciprocal equations by a simple substitution of $-V$ for V. This requirement is essential for the equivalence of both frames.

(ii) The transformation must convert any points in one frame located at a finite distance from the origin into points that are also at finite distances from an arbitrary origin in the other frame.

The first requirement greatly restricts the possible form of the transformations. For example, the transformation functions cannot be quadratic, because the inversion of a quadratic function leads to an irrationality, the same as an inversion of a function of any power but the first. A linear-fractional transformation, that is, the quotient of two linear expressions, can, under certain limitations imposed by the coefficients, be inverted with the help of a function of the same form. For example, for one variable, the direct and inverse linear-fractional functions look like this:

$$x' = \frac{ax+b}{ex+f}, \qquad x = \frac{b-fx'}{ex'-a}$$

But these functions do not satisfy the second condition: if $x' =$ $= a/e$, then x becomes infinite. Therefore only a linear function is possible.

(iii) When the relative velocity of the two frames tends to zero, the transformation equations yield an identity: $x = x'$, $y = y'$, $z = z'$, and $t = t'$.

(iv) The transformation equations yield a law of addition of velocities that leaves the velocity of light in vacuum invariant.

Summing up, we state briefly that the transformation equations (i) retain their form when inverted, (ii) are linear, (iii) become identities for small relative velocities, and (iv) leave the velocity of light in vacuum unchanged.

These four conditions are sufficient. To simplify the calculations, let us direct any two coordinate axes, say x and x', along the relative velocities of both systems. Then the coordinates along the other axes will not be affected by the transformation, that is, $y = y'$, $z = z'$.

Returning to Figure 9, we shall not make the arbitrary assumption that $t = t'$ (it is experimentally confirmed only at small relative velocities of the systems). Then the linear transformations of x and t can, in the most general form, be expressed as follows:

$$x' = \alpha x + \beta t \tag{13.2}$$

$$t' = \gamma x + \delta t \tag{13.3}$$

The coefficients α, β, γ, δ are determined from conditions (i)-(iv). There is no need to write the constant terms in these equations: they can be included in x or x' by appropriate choice of the origin of the reference frame.

Let us apply Eq. (13.2) to the origin of the primed frame of reference, $x' = 0$. This point moves with velocity V relative to the unprimed frame. Hence, $x = Vt$. Substituting $x' = 0$, $x = Vt$ into (13.2), we obtain, after eliminating t,

$$\alpha V + \beta = 0 \tag{13.4}$$

We shall solve equations (13.2) and (13.3) with respect to x and t. Elementary algebraic computations give

$$x = \frac{\delta x' - \beta t'}{\alpha \delta - \beta \gamma} \tag{13.5}$$

$$t = \frac{\gamma x' - \alpha t'}{\beta \gamma - \alpha \delta} \tag{13.6}$$

Let us now apply condition (i). For this we note that the coefficients β and γ, which interrelate the coordinate and time, must change sign together with the velocity V. Otherwise, if the x and x' axes are turned in the opposite direction, the equations will not preserve their form, and this is impermissible. A transformation

from x to $-x$ and from x' to $-x'$ is equivalent to a transformation from V to $-V$; for Eq. (13.2) to remain unchanged we must require that the sign of β reverses together with that of V. This also agrees with (13.4).

Thus, it is necessary for the equations of the inverse transformation from x' to x to differ from the direct transformation, (13.2) and (13.3), in the signs of β and γ:

$$x = \alpha x' - \beta t' \qquad (13.7)$$

$$t = -\gamma x' + \delta t' \qquad (13.8)$$

Comparing (13.7) and (13.5), we obtain

$$\alpha = \frac{\delta}{\alpha\delta - \beta\gamma} \qquad (13.9)$$

$$-\beta = \frac{-\beta}{\alpha\delta - \beta\gamma} \qquad (13.10)$$

From (13.10) it follows that

$$\alpha\delta - \beta\gamma = 1 \qquad (13.11)$$

Then (13.9) yields

$$\alpha = \delta \qquad (13.12)$$

This is all that is necessary for symmetry between the direct and inverse equations.

We now use condition (iv). We divide Eq. (13.2) by (13.3):

$$\frac{x'}{t'} = \frac{\alpha x/t + \beta}{\gamma x/t + \delta} \qquad (13.13)$$

Let x be a point occupied by an electromagnetic signal emitted from the origin of the unprimed frame at an initial instant of time $t = 0$. Obviously, $x/t = c$. But in accordance with condition (iv), $x'/t' = c$. Hence

$$c = \frac{\alpha c + \beta}{\gamma c + \delta} \qquad (13.14)$$

We substitute the relations (13.4) and (13.12) into (13.14) in order to eliminate β and δ. There remains a relation between α and γ:

$$\gamma c^2 + \alpha c = \alpha c - \alpha V$$

whence

$$\gamma = -\alpha \frac{V}{c^2} \qquad (13.15)$$

Substituting (13.15), (13.4), and (13.12) into (13.11), we obtain the equation for α:

$$\alpha^2 \left(1 - \frac{V^2}{c^2}\right) = 1 \qquad (13.16)$$

In extracting the square root we must take the positive sign in accordance with condition (iii), because then (13.3) becomes $t' = t$ for a small relative velocity. (Otherwise we would obtain $t' = -t$, which is meaningless.)

Now expressing all the coefficients α, β, γ, and δ in accordance with Eqs. (13.16), (13.4), (13.15), and (13.12), and substituting into (13.2) and (13.3), we arrive at the required transformations:

$$x' = \frac{x - Vt}{(1 - V^2/c^2)^{1/2}} \qquad (13.17)$$

$$t' = \frac{t - Vx/c^2}{(1 - V^2/c^2)^{1/2}} \qquad (13.18)$$

These equations are called the *Lorentz transformations*. From (13.7) and (13.8) the inverse transformations are of the form

$$x = \frac{x' + Vt'}{(1 - V^2/c^2)^{1/2}} \qquad (13.19)$$

$$t = \frac{t' + Vx'/c^2}{(1 - V^2/c^2)^{1/2}} \qquad (13.20)$$

In order to explain the meaning of these equations we shall apply them to some special cases. Let a clock be situated at the origin $x' = 0$ of the primed frame. It indicates a time t'. Then, from Eq. (13.20), it follows that

$$t = \frac{t'}{(1 - V^2/c^2)^{1/2}} \qquad (13.21a)$$

We shall call the clock at rest relative to the observer's reference frame the observer's clock. From formula (13.21a) we see that an observer comparing his clock, showing time t, with the clock of another observer must always conclude that the latter's clock is slow, that is, that $t' < t$. If the clock is at rest at the origin of the unprimed frame of reference, that is, at point $x = 0$, the transformation formula is of the same form, since from (13.18)

$$t' = \frac{t}{(1 - V^2/c^2)^{1/2}} \qquad (13.21b)$$

This not only does not contradict (13.21a) but expresses precisely the same fact: a clock moving relative to a certain observer lags behind his clock.

In relativity theory there is no single universal time as there is in Newtonian mechanics. It is better to say that the absolute time of Newtonian mechanics is an approximate concept, valid only at small relative velocities of the clocks being compared. The absolute nature of Newtonian time has sometimes given cause to regard it as an a priori, logical category independent of the motion of matter. But it should be remembered that in Newtonian mechanics the approximate concept of absolute time does not lead to contradictions, because action at a distance is assumed to be instantaneous. In formula (13.18) it is sufficient to put $c = \infty$ to obtain $t = t'$. In Newtonian mechanics instantaneous action was transmitted at a distance by gravitational forces.

It is sometimes assumed that, knowing the velocity of light c, it is possible to introduce a correction to the readings of clocks in different inertial systems for the passage of time to be always the same. But that is just what Eqs. (13.21*a*) and (13.21*b*) do: they describe the passage of time in two reference frames after introducing the correction for the finiteness of the propagation time of light. *Time dilation*, as it is called, is fully reciprocal in these reference frames and cannot, therefore, be ascribed to some change in the properties of the clocks associated with motion. Time dilation is a purely kinematic effect.

The relativity of time does not imply rejection of its objective nature. It is objective for every reference frame, just as the direction of a plumb line, which is different at different points of the globe, is an objective reality for each of them. Yet there was a time when the vertical direction was thought to be absolute.

Also relative, we find, is the concept of length of a line segment. In order to determine the length of a moving body—a "measuring-rod"—the coordinates of its ends must be simultaneously plotted in a fixed reference frame. There is no other basically different method of measuring a moving measuring-rod at the disposal of a stationary observer, since otherwise he would have to bring it to a halt, that is, transfer it to his own reference system. He must make the "notches" of the ends of the moving measuring-rod simultaneously according to his clock, say at the same instant $t = 0$. The concept of the simultaneity of two operations performed in the same reference frame can be uniquely defined with the help of light signals. Indeed, observers at rest relative to one another can always synchronize their clocks with the help of a light signal, making the necessary correction for the propagation time.

Substituting $t = 0$ into (13.17), we obtain the expression for the length of the moving measuring-rod relative to a stationary one:

$$\Delta x' = \frac{\Delta x}{(1 - V^2/c^2)^{1/2}} \tag{13.22}$$

If the observers exchange roles and the one moving with the measuring-rod measures the "unprimed" observer's rod, a similar formula results, in which $\Delta x'$ occurs in the right-hand side, and Δx in the left-hand side. Both equations relating the lengths of the stationary and the moving measuring-rods express the same fact: the moving rod contracts relative to the stationary one.

Note also that if, in an imaginary experiment, we undertook to photograph two fast-moving objects, no length contraction would be observed, because the light beams from the ends reaching the camera are not emitted simultaneously. An analysis reveals that in a snapshot the moving body would appear foreshortened as if viewed at an angle. Of course, making a photograph of such fast-moving bodies is, at present, as speculative as viewing them with the unaided eye.

Addition of Velocities. We shall now find the equation for the composition of velocities arising from the Lorentz transformations. Differentiating (13.17) and (13.18) and dividing one by the other, we obtain

$$\frac{dx'}{dt'} \equiv v'_x = \frac{dx/dt - V}{1 - (V/c^2)\,(dx/dt)} \equiv \frac{v_x - V}{1 - V v_x/c^2} \tag{13.23}$$

Noting that $dy' = dy$ and $dz' = dz$, we have a transformation of the velocity components perpendicular to V:

$$\frac{dy'}{dt'} \equiv v'_y = \frac{(dy/dt)\,(1 - V^2/c^2)^{1/2}}{1 - (V/c^2)\,(dx/dt)} = \frac{v_y\,(1 - V^2/c^2)^{1/2}}{1 - V v_x/c^2}$$

$$v'_z = \frac{v_z\,(1 - V^2/c^2)^{1/2}}{1 - V v_x/c^2} \tag{13.24}$$

For small velocities, (13.23) and (13.24) become the ordinary equations for addition of velocities. This can be seen if we let c tend to infinity, that is, by putting $V/c = 0$.

It is easy to see that if $v = (v_x^2 + v_y^2 + v_z^2)^{1/2} = c$, then likewise $v' = c$, that is, the absolute value of the velocity of electromagnetic perturbations does not change in passing from one inertial frame to another. But the separate components of the velocity of light, which are less than c, may of course change, just as the direction of a light ray relative to different observers may vary, since there is no absolute direction in space.

In this connection let us consider the phenomenon of the *aberration of light*. Astronomical aberration, or the deflection of light, consists in the fact that the stars describe small ellipses in the sky in the course of the year. The origin of this phenomenon is easy to explain: the velocity of the earth in its annual motion combines differently with the velocity of the light emitted by a star (Figure 22). If the

velocity vector of the light from the star relative to the sun is ES, then the resultant direction of the velocity for one position of the earth is ET_1, and in half a year it is ET_2. These directions are projected on different points of the celestial sphere, so that in the course of the year the star describes a closed ellipse. In the direction perpendicular to the plane of the earth's orbit both axes of the ellipse are equal, and we have a circle, while in the plane of the orbit the
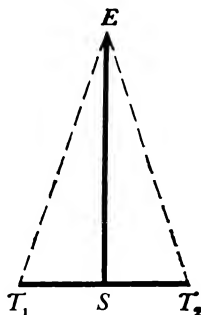

Figure 22

ellipse becomes a straight line equal in length to the diameter of that circle. In other words, the semi-major axis of the ellipse is always equal to V/c, where V is the velocity of the earth:

$$\frac{V}{c} = \frac{1}{10\,000} = 20.63''$$

But why, we may ask, doesn't the velocity of light in Michelson's experiment combine with the velocity of the earth and remains equal to c, while the phenomenon of aberration shows that these velocities do combine? To approximate the conditions of Michelson's experiment more closely with the observed aberration, the experiment was carried out with an extraterrestrial light source which, of course, did not change the result. The apparent contradiction is explained by the fact that in Michelson's experiment it is the absolute velocity of light that is measured according to the path difference of the beams, whereas in aberration the change in the direction of the velocity of light is due to the changing direction of the velocity of the earth along its orbit. If we take, for example, a star located in a direction perpendicular to the earth's orbit, then in (13.23) and (13.24) we have $v_x = 0$, $v_y = c$, $v_z = 0$. The components of the velocity of light relative to the earth are

$$v_x' = -V, \qquad v_y' = c\,(1 - V^2/c^2)^{1/2}$$

And, in accordance with Michelson's experiment, $v'^2_x + v'^2_y = c^2$. The direction of the projection of the velocity of light on the plane of the earth's orbit (the ecliptic) is reversed in the course of half a year, which is why aberration occurs.

Similar equations are obtained in the more complicated case when the light rays are not perpendicular to the plane of the ecliptic. They coincide fully with the conventional equations for the addition of velocities only when terms of the order V^2/c^2 are neglected.

Another contradiction with Michelson's experiment was thought to be demonstrated in the Fizeau experiment, designed to determine
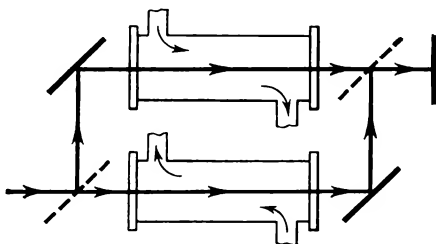
Figure 23

the velocity of light relative to a moving medium. The Fizeau method was this. A beam of light was divided into two parts using a half-silvered mirror (Figure 23). These beams were passed through tubes with flowing water, one beam in the direction of flow and the other in the opposite direction. For comparison, the same beams were passed through the tubes with the water at rest. By subsequent reflections the beams once again combined and cancelled each other when the path difference between them was equal to an integral number of half wavelengths (that is, when they were in opposite phase). Coherence between them was obtained due to the fact that they both came from the same source. In still water, the path difference was chosen so that the rays were reinforced, that is, the phase difference was equal to an integral number of wavelengths. The path difference in flowing water was varied. Since the frequency of the light and the tube lengths remained unchanged, the change in path length indicated that the speed of light in flowing water differed from that in still water.

First of all, we note that the result of the Fizeau experiment in no way contradicts the general ideas about the relativity of motion. A reference system fixed in flowing water is not equivalent to a system fixed in the tube, if we are studying the propagation of light in water.

Since the velocity of light in water is equal to c/n, where n is the refractive index of the water, the general equation for the addition of velocities (13.23) shows that c/n does not remain a constant quantity when passing to another reference frame. At the same time we cannot use the simple velocity-addition equation, because the denominator of Eq. (13.23) differs from unity by $V/(nc)$, that is, a quantity of the same order as the quantity in parentheses in the numerator, if it is represented as $(c/n)\,(1 - Vn/c)$, where V is the velocity of the water. Assuming that $V \ll c$, and expanding the denominator in a series up to the linear term inclusive, we find the change in the velocity of light in flowing water (see Exercise 1):

$$u = \frac{c}{n} \pm V\left(1| - \frac{1}{n^2}\right)$$

This does not quite coincide with the result that could be expected if we applied the conventional formula for the composition of velocities. When Fizeau performed his experiment (in the mid-nineteenth century), the result proved to be somewhat unexpected. Relativity theory explained the appearance of the factor $(1 - 1/n^2)$. And since Michelson measured the quantity c, and Fizeau the quantity c/n, there is no contradiction between their experiments. It should also be noted that Michelson's experiment detected a quadratic effect (or rather its absence!), while the Fizeau experiment dealt with an effect linear with respect to V. That is why in Michelson's experiment the earth's velocity of 30 km·s^{-1}, which is much greater than the speed of water flowing through a pipe, is used.

The Interval. Despite the fact that x and t are changed separately by the Lorentz transformations, we can construct a quantity which remains invariant (unchanged). It is easy to verify that this property is possessed by the difference $c^2t^2 - x^2$. Indeed

$$c^2t^2 = \frac{c^2t'^2 + V^2x'^2/c^2 + 2Vx't'}{1 - V^2/c^2}$$

$$x^2 = \frac{x'^2 + V^2t'^2 + 2Vx't'}{1 - V^2/c^2}$$

or

$$c^2t^2 - x^2 = c^2t'^2 - x'^2 \equiv s^2 \tag{13.25}$$

The quantity s is called the *interval between two events*: that which occurred at the coordinate origin $x = 0$ at the initial time $t = 0$, and another event that occurred at the point x at time t.

The word "event" may be considered in its most common everyday sense, provided that its coordinates and time may be defined. If the first event is not related to the origin of the coordinate system

and the initial instant, then

$$s^2 = c^2 (t_2 - t_1)^2 - (x_2 - x_1)^2 = c^2 (t_2' - t_1')^2 - (x_2' - x_1')^2 \quad (13.26)$$

Considerable importance is attached to the interval between two infinitely close events. We shall assume that they are separated by a length segment oriented arbitrarily rather than with respect to the x axis. Then an infinitesimal interval ds between the two events is defined as follows:

$$\begin{aligned} ds^2 &= c^2 \, dt^2 - dx^2 - dy^2 - dz^2 \\ &= c^2 \, dt'^2 - dx'^2 - dy'^2 - dz'^2 = c^2 \, dt'^2 - dl'^2 \end{aligned} \quad (13.27)$$

The interval thus written is not related to any definite direction of the relative velocity of the reference frames.

Spacelike and Timelike Intervals. The interval concept provides an extremely vivid method for studying various possible space-time relationships between two events. Let the spatial distance between
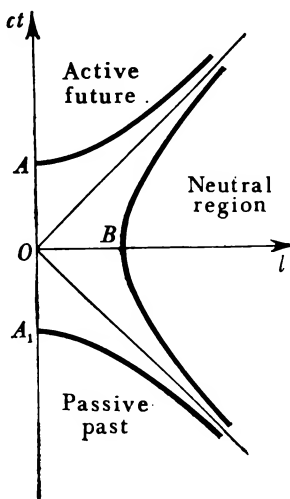
Figure 24

the points in which the events occurred be taken along the abscissa, and let the time interval between them be plotted along the ordinate (Figure 24).

Imagine the case of $ct > l$, i. e. $s^2 > 0$. If the same two events are considered in different inertial reference frames, the time intervals and spatial distances between them will be quite different, but the interval $s^2 = c^2 t^2 - l^2$ remains invariant in all reference

frames. It follows that the locus of all possible spatial distances l^2 and time intervals $(ct)^2$ is an equilateral hyperbola, $s^2 = c^2t^2 - l^2$. One branch of the hyperbola lies in the past relative to the event that occurred at time $t = 0$, $x = 0$, while the other is completely in the future. It is easy to see that such a relationship is inevitable if the events are causally related. Let two events have occurred at the same spatial point in some reference system, the second event being an effect of the first. To this reference frame there correspond a point O (cause) and a point A (effect). But all the points of the hyperbola through A lie at $t > 0$, so that the cause precedes the effect in any reference frame.

We may also proceed from causally related events taking place at different spatial points in the initial reference frame, for example the firing of a shot and the hitting of a target. But in the frame fixed relative to the bullet both events are located on line OA (Figure 24). In the initial frame of reference the firing of the bullet and the hitting of the target lie on an inclined line drawn from the origin to the same hyperbola passing through A. Therefore in any reference frame the target is hit after the shot.

If we denote the velocity of the bullet (or of any particle) v, then it is easy to see that $v < c$. Indeed, for a reference frame fixed relative to the particle to exist we must assume that $ds^2 \geqslant c^2\, dt^2 > 0$. But then $dl^2 = v^2\, dt^2 < c^2\, dt^2$, or $v < c$, as was asserted. The velocity of light is the limiting velocity for a material particle.

The domain above the first asymptote is called the *active future* with respect to the initial event. If the cause is an event at point O, then every effect lies in that domain. Thus, the theory of relativity does not contradict the objective nature of causality.

Other examples of spatio-temporal configurations of events can be offered, for which $c^2t^2 < l^2$ and $s^2 < 0$. Such events can in no way be causally related. As we have seen, the speed with which matter is transported cannot exceed c, while to have $s^2 < 0$ we must put $l^2 > c^2t^2$. There does not exist a reference frame in which both these events could occur at the same spatial point. For them $s^2 < 0$, so that the interval is an imaginary quantity.

However, the time sequence of such events is not defined: there are reference frames in which the event arbitrarily described as the first occurred before the second event, and reference frames in which the second event occurred before the first. Thus relativity theory denies the absolute nature of the simultaneity of two events the interval between which is imaginary. There is no reference frame in which they could have occurred at the same point in space. An example are points O and B in Figure 24. Point B lies on a hyperbola belonging partly to the future and partly to the past. It is clear that O and B cannot be causally related, since otherwise the interaction would have to propagate from O to B instantaneously. Hence

there exist reference frames in which event B occurred before O.

The domain between the asymptotes on the side of B is called the *neutral region* with respect to the initial event.

The asymptotes themselves are of special interest: on them $l = ct$, $s = 0$. The relationship $l = ct$ occurs in the case of two events connected by an electromagnetic signal, for example the dispatch and reception of a wireless message. For two such events, $s = 0$ in all reference frames, because the speed of light is invariant, and we must always have $l = ct$. Since the graph in Figure 24 actually refers not to a plane but to $3 + 1$ dimensions (the three spatial coordinates and time), the locus of zero intervals is described as the *light cone*.

Proper Time. Closely related to the interval concept is that of the *proper time* of a particle. By definition, the displacement of a particle relative to a reference frame fixed with respect to it is zero. The reference frame is not necessarily inertial if the particle is moving with acceleration. Time measured in the particle's own system is, apparently, expressed in terms of the interval as

$$dt_0 = \frac{ds}{c} = \left(dt^2 - \frac{dl^2}{c^2} \right)^{1/2} = dt \left(1 - \frac{v^2}{c^2} \right)^{1/2} \tag{13.28}$$

Here, dt and dl represent the time interval and displacement of the particle relative to a reference frame not associated with it. From (13.28) or (13.21), the proper time of a particle travelling with velocity $v = dl/dt$ is the shortest. For finite time intervals we obtain

$$t_0 = \int dt \left(1 - \frac{v^2(t)}{c^2} \right)^{1/2} \leqslant \int dt \tag{13.29}$$

or

$$t_0 \leqslant t \tag{13.30}$$

It may seem that (13.29) and (13.30) contradict what was said concerning the reciprocal nature of time dilation for two observers. Actually, to measure the difference in times yielded by these equations the moving reference frame must be brought to rest relative to the stationary one, that is, it must be deprived of its inertial character. Such a noninertial frame is in no way equivalent to an inertial one.

The foregoing can also be explained by means of the following reasoning. Let a reference frame be accelerated during a certain time interval τ; then it travels uniformly for a very long time t; then its direction of motion is reversed during time 2τ; then it again travels uniformly for a long time t; and finally it slows down to rest in time τ. This, of course, is a description of a round trip.

According to Eq. (13.29), the time dilation occurs mainly during the period of uniform motion, since $t \gg \tau$. If that period is increased,

say, tenfold, while the acceleration and deceleration times remain the same, the time dilation is also tenfold greater. It follows from this that the acceleration and deceleration are needed only to com-- pare the passage of time in the two reference frames, the inertial and the noninertial, but not to effect time dilation in the noniner- tial frame. Noninertionality is, so to say, a tag on the system that makes it possible to single it out as the one in which less time was seen to have passed between the departure and return from the trip.

Mathematically this can also be explained as follows: dt is a total differential, while $dt_0 = ds/c$ is not a total differential. Therefore the value of the integral (13.29) depends on the function $v(t)$ sub- stituted into it. We can discern here an analogue between the lengths of a broken or a curved line and the difference between the coordi- nates of its end points. Coordinates are a definite quantity, while the length of a path depends upon the form of the curve.

Time t_0 is the reading of the clock in the moving reference frame. It defines the rhythm of all physical (and physiological) processes in that frame. This can at present be directly verified experimentally only with regard to the decay time of elementary particles, but even without direct verification there is no doubt at all about the validity of the conclusion.

The mean lifetime of a positive π-meson, whose mass is equal to 276 electron masses, before it decays into a μ-meson, or muon, with a mass of 206 electron masses, and a neutral particle, is 2×10^{-8} s (negative π-mesons are usually captured by nuclei before they have time to decay). This time has been measured for π-mesons brought to rest in matter, that is, it is their proper time. If the relationship (13.30), which expresses time dilation, did not exist, a fast π-meson would have the same lifetime relative to a stationary reference frame fixed with respect to the earth. In that case it would be able to travel no more than $2 \times 10^{-8} \times 3 \times 10^{10} = 600$ cm through air, because c is the limiting velocity of motion. Actually the mean path of π-me- sons is much longer, because their lifetime measured in a stationary reference frame can be much longer than in their proper reference frame.

Tensor Notation. In Section 11 it was pointed out that, besides possessing the same dimensionality on both sides, a correct physical equation must also be invariant with respect to rotations of the coordinate system. In other words, only such quantities may occur on both sides of the equation which transform similarly in passing to another coordinate system. For this to be apparent from the equation itself it is convenient to write it in vector or tensor form. Furthermore, every equation must satisfy the relativity principle,

that is, preserve its form in transformations to other inertial reference brames.

As applied to Einstein's relativity principle, it is possible to write equations in a way that reveals both the necessary properties of invariance simultaneously. Such notation is known as *relativistic invariant notation.*

First we give a somewhat different form to the Lorentz transformations, introducing the notation

$$i\,\frac{V}{c} \equiv \tan\psi \tag{13.31}$$

so that

$$\left(1 - \frac{V^2}{c^2}\right)^{1/2} = \frac{1}{\cos\psi}$$

We further introduce the imaginary coordinate x_4

$$x_4 = ict \tag{13.32}$$

Then transformations (13.17) and (13.18) take the form of a rotation of the coordinate system through the angle ψ:

$$x_1' = x_1 \cos\psi + x_4 \sin\psi \tag{13.33}$$

$$x_4' = -x_1 \sin\psi + x_4 \cos\psi \tag{13.34}$$

This manipulation involves no additional physical assumptions in comparison with those made in developing the set of transformations (13.17) and (13.18). As will now be shown, the improvements were made only in the notation. The imaginary unit is needed to achieve complete formal similarity with conventional rotation of coordinates.

Any rotation of coordinates, including conventional (spatial) rotation, can be pictured as a set of separate rotations in which only two of the total number of coordinates are transformed. In particular, if a fourth coordinate is introduced, then any rotation in four-dimensional space is effected by the Lorentz transformations in the form (13.33) and (13.34) and by additional spatial rotations of the coordinates through real angles. We did not make time an imaginary coordinate, but by multiplying it by an imaginary unit we have been able to transform the set of three spatial coordinates and time as a single four-dimensional manifold of Cartesian coordinates.

It is natural here to introduce four-dimensional scalars, vectors, and tensors. For example, the interval immediately displays its scalarly invariant nature. Introducing into it $c\,dt = dx_4/i$, we obtain

$$ds^2 = c^2 dt^2 - dx_1^2 - dx_2^2 - dx_3^2 = -dx_k\,dx_k = -(dx_k)^2 \tag{13.35}$$

Here the dummy subscript k takes on values from 1 to 4, as should be in a scalar expression. Unlike the indices in three-dimensional

space, which in tensor summation assume values from 1 to 3 and are denoted by Greek letters, we shall denote four-dimensional space indices by Latin letters.

The known definition of a vector remains: it is a collection of four quantities which transform like coordinates. We shall encounter such quantities in relativistic mechanics (Section 14) and electrodynamics (Section 15).

All the main results referring to tensors in three-dimensional space are applicable to four-dimensional space, with the exception of one more or less "fortuitous" one, the vector product written in three-dimensional space as

$$(\mathbf{A} \times \mathbf{B})_\alpha = \varepsilon_{\alpha\beta\gamma} A_\beta B_\gamma$$

Having to deal essentially with an antisymmetric tensor of rank 2, $A_\beta B_\gamma - A_\gamma B_\beta$, we reduced it to a vector, since the number of components of the one and the other in three-dimensional space is the same. In four-dimensional space the antisymmetric tensor A_{ik} has six components, A_{12}, A_{13}, A_{14}, A_{23}, A_{24}, A_{34}, while a vector has only four, and there can be no correspondence between them. That is why the operation of vector multiplication, wherein lies the basic meaning of vector analysis, does not occur in four-dimensional space. Here, tensor notation is doubtlessly more convenient.

We shall prove the relativistic invariance of an equation by reducing it to four-dimensional tensor form. This form of notation, in turn, makes it possible to specify in advance the equations that agree with relativity theory, and thus it substantially reduces the number of possible physical assumptions in the search for new laws and regularities.

EXERCISES

1. Calculate the change in the velocity of light propagating through flowing water in the Fizeau experiment.

Solution.

$$u_\pm = \frac{c/n \pm V}{1 \pm V/nc} \approx \left(\frac{c}{n} \pm V\right)\left(1 \mp \frac{V}{nc}\right) \approx \frac{c}{n} \pm V\left(1 - \frac{1}{n^2}\right)$$

Disregarding the theory of relativity, the result would be $u_\pm = c/n \pm V$.

2. Obtain a precise equation for the aberration of light, for an arbitrary inclination of the ray to the ecliptic.

Answer.

$$\cos \vartheta' = \frac{\cos \vartheta - V/c}{1 - V/c \cos \vartheta}$$

3. Write the equations for the Lorentz transformations for an arbitrary direction of the velocity **V** relative to a coordinate system.

Solution. In our equations $x = \mathbf{rV}/V$ and $x' = \mathbf{r'V}/V$. The radius-vector component perpendicular to the velocity is

$$\mathbf{r} - \frac{\mathbf{V}\,(\mathbf{r}\cdot\mathbf{V})}{V^2} = \mathbf{r'} - \frac{\mathbf{V}\,(\mathbf{r'}\cdot\mathbf{V})}{V^2}$$

From (13.17)

$$\frac{\mathbf{r'V}}{V} = \frac{\mathbf{rV}/V - Vt}{(1 - V^2/c^2)^{1/2}}$$

Multiplying this equation by **V**/V and adding the equations, we obtain

$$\mathbf{r'} = \mathbf{r} - \frac{\mathbf{V}\,(\mathbf{r}\cdot\mathbf{V})}{V^2} + \left(\frac{\mathbf{V}\,(\mathbf{r}\cdot\mathbf{V})}{V^2} - \mathbf{V}t\right)\left(1 - \frac{V^2}{c^2}\right)^{-1/2}$$

4. Show that a four-dimensional element of volume, $dx_1 dx_2 dx_3 dx_4$, is invariant under the Lorentz transformations.

Solution. Represent the volume element in tensor notation:

$$d^{(4)}\tau = \varepsilon_{iklm}\, dx_i^{(1)}\, dx_k^{(2)}\, dx_l^{(3)}\, dx_m^{(4)}$$

where only the ith component of a vector, dx_i, is other than zero.

We may also write the array of transformation coefficients corresponding to the transformation from the unprimed to the primed reference frame according to (13.17) and (13.18):

$$\begin{pmatrix} \dfrac{1}{\sqrt{1 - V^2/c^2}} & 0 & 0 & \dfrac{-V/c^2}{\sqrt{1 - V^2/c^2}} \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -V & 0 & 0 & \dfrac{1}{\sqrt{1 - V^2/c^2}} \end{pmatrix}$$

The Jacobian of this transformation is unity. A transformation of more general form is associated with subsequent rotations in three-dimensional space which do not affect the corresponding three-dimensional volume element $d^{(3)}\tau$. Hence, $d^{(4)}\tau$ is always invariant.

5. Find how the Lorentz transformations alter a three-dimensional volume element.

Solution. The Jacobian comprising the first three rows and columns of the transformation array is equal to $(1 - V^2/c^2)^{-1/2}$, whence

$$d^{(3)}\tau' = d^{(3)}\tau\,\frac{1}{(1 - V^2/c^2)^{1/2}}$$

This can also be demonstrated by applying the length contraction formula (13.22) to the volume element in the direction parallel to the velocity. If t from (13.21) is substituted, the product $d^{(3)}\tau dt$ again yields an invariant.

14

RELATIVISTIC MECHANICS

The concept of relativistic invariance will enable us to obtain the expression for the action of a free mass point. Much of what was said concerning the action function in Newtonian mechanics remains valid in relativity theory; specifically action must not involve the coordinates and time explicitly under the integral sign, and it cannot be dependent upon the direction of the particle's velocity. It must also satisfy the requirement of the relativity principle that its form must not change under transformations to other inertial frames. It was pointed out in Section 13 that when the equations are written in a relativistic invariant form, the invariance condition relative to spatial rotations is satisfied together with Einstein's relativity principle.

Another necessary requirement which action must satisfy in relativity theory is that, at small velocities, all expressions become the corresponding expressions of Newtonian mechanics.

The Lagrangian of a Free Particle. In Section 13 we obtained an infinitesimal of the first order with respect to the increment of all spatial variables of a particle and to the time increment. This is the infinitesimal interval ds, which also satisfies the requirement of relativistic invariance. No other such quantity can be developed. We shall therefore look for the action of a free particle in the form

$$S = \alpha \int ds \tag{14.1}$$

We now pass from the action to the Lagrangian. For this we represent the infinitesimal interval as follows:

$$ds = (c^2\, dt^2 - dl^2)^{1/2} = c\, dt \left[1 - \left(\frac{1}{c} \frac{dl}{dt} \right)^2 \right]^{1/2}$$
$$= c\, dt\, (1 - v^2/c^2)^{1/2} \tag{14.2}$$

Hence, the Lagrangian, which is identically determined by $S \equiv$ $\equiv \int L\, dt$, is

$$L = \alpha c\, (1 - v^2/c^2)^{1/2} \tag{14.3}$$

We determine the coefficient α from the condition that, at small velocities v, the function L turns into the nonrelativistic expression of the action of a free particle. Since

$$\left(1 - \frac{v^2}{c^2} \right)^{1/2} \approx 1 - \frac{v^2}{2c^2}$$

for $v \ll c$ we rewrite L in the form

$$L = \alpha c \left(1 - \frac{v^2}{c^2} \right)^{1/2} = \alpha c - \frac{c\alpha v^2}{2c^2}$$

Since the first term is constant, it can be omitted, while the second term should reduce to $mv^2/2$ (see Sec. 2). Comparing with (2.26), we obtain

$$\alpha = -mc \tag{14.4}$$

According to the meaning, here the mass m of the particle is defined in its proper reference frame. In future this is the only definition of mass we shall use. Since the proper reference frame is stated uniquely, the quantity m is relativistically invariant and characterizes the particle. Finally, we have the Lagrangian in the form

$$L = -mc^2 \left(1 - \frac{v^2}{c^2} \right)^{1/2} \tag{14.5}$$

Momentum in Relativistic Mechanics. The expression for momentum in relativity theory is obtained directly from (14.5):

$$\mathbf{p} = \frac{\partial L}{\partial \mathbf{v}} = \frac{m\mathbf{v}}{(1 - v^2/c^2)^{1/2}} \tag{14.6}$$

At small velocities it reduces, as it should, to the expression for momentum in Newtonian mechanics, $\mathbf{p} = m\mathbf{v}$.

In some books the quantity $m(1 - v^2/c^2)^{-1/2}$, that is, the proportionality factor between velocity and momentum, is called the "mass of motion", as distinct from the *rest mass*, m. To avoid confusion, we shall not use the term "mass of motion", and will always take the term "mass" to mean the relativistically invariant quantity m.

The limiting nature of the velocity of light, mentioned before in Section 13, can be seen from Eq. (14.6). As a particle's velocity tends to the speed of light its momentum tends to infinity.

The only possible exception is a particle of mass zero. The momentum of such a particle written in the form (14.6) yields, for $v = c$, an indeterminacy of the form $0/0$ and may remain finite. But then the velocity of such a particle must be equal to c, since otherwise its momentum vanishes identically, making it incapable of interacting with any mechanical system, that is, it would in no way display its physical existence.

As we know, the velocity c is relativistically invariant, so that the property of a given particle to travel with the speed of light is inherent in it, not in the reference frame in which its motion is described. The momentum of such a particle must be stated not according to Eq. (14.6) but independently of the magnitude of its velocity, which is always the same and equal to c.

Velocities greater than c are physically meaningless, since they would be associated with imaginary values of the momentum. Particles travelling faster than light would be moving faster than interactions transmitted between them. The absurdity of such a situation can be readily appreciated: the causality principle would break down. The impossibility of velocities exceeding the speed of light stems from the fact that the relativity principle does not violate causality, which links objective events. And in Section 13 it was already pointed out that such a concept as the order of cause and effect does not depend on the choice of reference frame.

Energy in the Theory of Relativity. We proceed from the general definition of energy (4.1):

$$E= \mathbf{v}\,\frac{\partial L}{\partial \mathbf{v}} - L = \frac{mv^2}{(1-v^2/c^2)^{1/2}} + mc^2\,(1-v^2/c^2)^{1/2}$$
$$= \frac{mc^2}{(1-v^2/c^2)^{1/2}} \tag{14.7}$$

This equation reaffirms the limiting nature of the velocity of light. When $v \to c$, the energy of particle tends to infinity. In other words, infinite work must be done to accelerate a particle to the speed of light. An exception are particles whose mass is zero, and at $v = c$ they have finite momentum and finite energy.

From Eq. (14.7), the energy of a particle at rest is equal to mc^2. Let us apply this formula to a compound particle capable of spontaneously decaying into two new particles, for example, a nucleus decaying into a daughter nucleus and an alpha-particle. Since the decay is spontaneous, it is due not to an external action on the parent nucleus but to certain specific features of its internal motion. Consequently, radioactive decay is a process that takes place in a closed system, and the total energy is therefore conserved. The energy of the paternal particle prior to disintegration is equal to the sum of the energies of the product nucleus and the α-particle after disintegration, when they no longer interact.

The energy of each particle is expressed according to equation (14.7), which is applicable to any particle (simple or compound) whose motion is considered as a whole. The only possible form of the Lagrangian for such motion is (14.5), whence it follows that the energy is represented in the form (14.7). Assuming now that the decaying particle was at rest, we write the expression for the energy-conservation law in the decay process, using (14.7):

$$mc^2 = \frac{m_1c^2}{(1-v_1^2/c^2)^{1/2}} + \frac{m_2c^2}{(1-v_2^2/c^2)^{1/2}} \tag{14.8}$$

Both terms in the right-hand side, E_1 and E_2, are respectively greater than m_1c^2 and m_2c^2, whence we obtain the important

inequality

$$m > m_1 + m_2 \tag{14.9}$$

Thus, the mass of a compound particle capable of spontaneous decay is greater than the sum of the masses of the particles created as a result of the disintegration. This is an essentially new fact of relativistic mechanics as compared with Newtonian mechanics, where the law of additivity of mass holds. If we define the difference

$$T \equiv \frac{mc^2}{(1 - v^2/c^2)^{1/2}} - mc^2 \tag{14.10}$$

as the kinetic energy of a particle (for small energies it reduces to $mv^2/2$) and call mc^2 the *rest energy*, then it can be seen from the law of conservation of energy (14.8) that part of the rest energy of a compound particle is converted into the kinetic energy of the component particles, and part is converted into their rest energy. Only the total energies E, and not the kinetic energies T, satisfy the conservation law, because the kinetic energy of the parent particle as a whole is equal to zero before disintegration and cannot be equal to the essentially positive kinetic energy of the decay products.

In chemical reactions, the change in the rest masses of the reacting substances occurs in the order of 10^{-9} (and less) of the total mass. In nuclear reactions, where the particle velocities are of the order $c/10$, the change in mass may approach one-half of one per cent. When an electron and positron (a positive electron) are annihilated, their energy, including rest energy, is totally converted into the energy of electromagnetic radiation.

From the quantum theory (see Part III) it is known that radiation propagates through space in the form of separate particles, called light quanta. This is not only compatible with the wave properties of radiation but derives directly from them. The velocity of a light quantum is c, so that its mass is identically equal to zero. The total rest mass of the particles taking part in an annihilation process is equal to $2mc^2$ before annihilation, and to zero afterwards. The increase in the energy of the electromagnetic field is, of course, not less than $2mc^2$. The least value $(2mc^2)$ is obtained when an electron and positron annihilate at rest and have no additional kinetic energy. We could call the energy of an electromagnetic field divided by c^2 its "mass". With such a definition of mass, the total "mass" would be conserved. But such a "mass" conservation contains nothing new in comparison with the law of conservation of energy. Dividing the equation expressing the latter by c^2 yields no essentially new law; all that occurs is a transformation to other measurement units.

It is precisely the rest mass that is best used in determining the energy balance in nuclear reactions, for a change in the rest mass of all the particles involved in the transmutation determines the

energy which may be generated as a result of the reaction in the form of the kinetic energy of the disintegration products or radiation energy. There is no sense in calling the energy of a light quantum divided by the square of the velocity of light its mass, because this quantity does not in any way characterize light quanta. Energy has one value in one reference system and an entirely different value in another, whereas mass is a quantity that characterizes a particle. For example, the mass of an electron (the rest mass) is equal to 9×10^{-28} g, while the corresponding quantity for a quantum is identically equal to zero. But this zero characterizes a light quantum to no less an extent than 9×10^{-28} g characterizes an electron.

The mass of a particle determines the relationship between momentum and velocity according to Eq. (14.6). It is impossible to calculate a particle's mass from its momentum alone, because particles with the same momenta may have quite different masses. That is why the assertion occasionally made that the fact that electromagnetic field possesses momentum, which is manifested in the form of light pressure, is proof that light quanta possess mass is quite meaningless. Mass cannot be determined from momentum or energy separately: it is involved only in the relationship between the two quantities for the given particle and thereby characterizes that particle.

Also erroneous is the widespread assertion that a mass of one gram is capable of evolving 9×10^{20} ergs of energy. For that, one-half of the mass would have to be antimatter and annihilate together with the matter. Changes in mass in pure matter (or pure antimatter) occur in nuclear reactions, where the total number of protons and neutrons does not change. Therefore, the change in mass amounts to no more than fractions of a percentage point.

We shall now express energy in terms of momentum. Squaring Eq. (14.7) and subtracting from it Eq. (14.6), after the latter has also been squared and multiplied by c^2, we obtain

$$E^2 - c^2 p^2 = m^2 c^4 \tag{14.11}$$

In Section 10 we have called the energy expressed in terms of momentum the Hamiltonian. Hence

$$E \equiv \mathcal{H} = (m^2 c^4 + c^2 p^2)^{1/2} \tag{14.12}$$

Whence we obtain a relationship between the energy and momentum of a particle that has no mass:

$$E = cp \tag{14.13}$$

This is the form to which expression (14.12) tends when the momentum tends to infinity.

The Lorentz Transformations for Momentum and Energy. To keep track of the relativistic invariance of equations it is convenient to

write them in four-dimensional tensor notation. Let us first represent momentum and energy in such notation.

We substitute $dt/ds = c^{-1}(1 - v^2/c^2)^{-1/2}$ into the definition of momentum (14.6). Its components then take the form

$$p_\alpha = mc \frac{dx_\alpha}{ds} \tag{14.14a}$$

where the Greek index α assumes values from 1 to 3, as agreed.

In the definition of energy (14.7), velocity must be written as dl/dt, and dt replaced by $dx_4/(ic)$. Comparison with (14.14a) then shows that the imaginary quantity iE/c should be treated as the fourth component of the momentum vector:

$$p_4 = mc \frac{dx_4}{ds} \tag{14.14b}$$

Together, the momentum and energy of a particle form a single four-vector with components

$$p_i = mc \frac{dx_i}{ds} \tag{14.15}$$

But it was shown in Section 13 that from the mathematical point of view a Lorentz transformation represents a rotation of a reference frame through an imaginary angle the tangent of which is equal to iV/c. By definition, every vector transforms as a radius vector. Hence, in passing to another inertial reference frame the components of a four-momentum must transform according to Eqs. (13.33) and (13.34):

$$p_1' = p_1 \cos \psi + p_4 \sin \psi$$
$$p_4' = -p_1 \sin \psi + p_4 \cos \psi \tag{14.16}$$

In order to return to conventional three-dimensional notation we substitute $\tan \psi = iV/c$ and $p_4 = iE/c$. We then find the required momentum and energy transformation formulas:

$$p_x' = \frac{p_x - EV/c^2}{(1 - V^2/c^2)^{1/2}} \tag{14.17}$$

$$E' = \frac{E - Vp_x}{(1 - V^2/c^2)^{1/2}} \tag{14.18}$$

If the relative velocity of the reference frames is directed along the x axis, p_y and p_z do not change.

Note that a correct limiting process from (14.17) to the nonrelativistic momentum transformation formula $p_x' = p_x - mV$ is obtained only if in place of E we substitute the rest energy mc^2. The nonrelativistic equation corresponds to a simple addition of velocities: $v_x' = v_x - V$.

Hence, if we demand that the Lorentz transformations yield the correct limiting transition to the Galilean transformations, it is necessary to include the rest energy of the particles in their total energy. Conversely, the kinetic energy T from (14.10) does not give a correct limiting transition.

In relativistically invariant form, the relationship (14.12) between energy and momentum is written as follows:

$$p_i p_i \equiv p_i^2 = -m^2 c^2 \tag{14.19}$$

It is based on the fact that the square of the four-dimensional velocity $u_i \equiv dx_i/ds$ is equal to -1:

$$u_i u_i = \frac{dx_i \, dx_i}{ds^2} = -1$$

The Velocity of a System of Noninteracting Particles in Relativity Theory. We shall now show how to determine the velocity of a system of noninteracting particles in the theory of relativity. The difference from Newtonian mechanics is that the description of an interaction requires the inclusion of the field in the system, and the particles can no longer be treated by themselves as a closed system, even if there are no external fields acting on them.

For simplicity's sake we consider two particles. Between the velocity, energy, and momentum of each particle there exists the relation

$$\mathbf{p} = \frac{E \mathbf{V}}{c^2} \tag{14.20}$$

which is derived from (14.6) and (14.7). The same equation can also be obtained somewhat differently. From (14.17) we determine the velocity of the reference frame in which the particle's momentum is zero. Putting $p_x' = 0$ in the left-hand side of (14.17), we have in the right-hand side $V = p_x c^2/E$, or, if the velocity is not directed along the x axis, in general $\mathbf{V} = \mathbf{p} c^2/E = \mathbf{v}$, in agreement with (14.20). Applied to one particle, the equality $\mathbf{V} = \mathbf{v}$ is trivial and denotes simply that the momentum of a particle relative to a reference frame moving with the same velocity is zero.

We now apply (14.20) to the two particles so as to find the velocity of the reference frame relative to which their total momentum is zero. Since the particles do not interact, their total momentum and total energy are simply added; otherwise we would have to take into account the momentum and energy of their interaction field (in the case of nuclear forces we simply don't know how to do this).

And so, the total momentum of the particles is $\mathbf{p}_1 + \mathbf{p}_2 = \mathbf{p}$, and their total energy is $E_1 + E_2 = E$. We direct the x axis along $\mathbf{p}$. Since the Lorentz transformations are linear and homogeneous, the transformation formulas to another inertial frame for a sum of two

four-vectors look the same as for each of them separately. Hence, the required velocity of the reference frame in which the total momentum is zero is given as

$$\mathbf{V} = \frac{c^2 (\mathbf{p_1} + \mathbf{p_2})}{E_1 + E_2} \tag{14.21}$$

To obtain the limiting transition to Newtonian mechanics from this we must put $\mathbf{p_1} = m_1 \mathbf{v_1}$, $\mathbf{p_2} = m_2 \mathbf{v_2}$, $E_1 = m_1 c^2$, $E_2 = m_2 c^2$. Then Eq. (14.21) coincides with the conventional expression for the velocity of the centre of mass cf a system of particles. The quantity $\mathbf{V}$, expressed in terms of the particles' velocities according to (14.21), does not have the form of a total time derivative of any coordinate, hence it is impossible to determine the coordinates of the centre of mass from its velocity.

In relativistic mechanics the relative velocity $\mathbf{v_1} - \mathbf{v_2}$ is meaningless, because there is no simple law of velocity addition.

Action for a Charged Particle in an External Electromagnetic Field. Let us now consider the equations of motion of a charged particle in an external electromagnetic field. As in the case of a free particle, we shall proceed from the expression for action, requiring that it be relativistically invariant.

Of course, one could write many possible expressions that would be relativistically invariant in form. We find, however, that one of the simplest expressions is in agreement with experience. By "experience" we mean a totality of facts at least as great as the foundation on which Newtonian mechanics rests.

The action of a particle in an electromagnetic field includes the action of a free particle and a supplementary term describing the interaction of the electromagnetic field and the charge; in relativistically invariant notation it looks like this:

$$S = \int \left(-mc \, ds + \frac{e}{c} A_k \, dx_k \right) \tag{14.22}$$

where A_k is meant to denote a four-vector. Its three spatial components yield the known vector potential of an electromagnetic field derived in Section 12. (This will be shown in the following section in developing Maxwell's equations from the relativistically invariant action of an electromagnetic field.) The fourth component, A_4, is $i\varphi$, where φ is the scalar potential also developed in Section 12. The constant e is called the charge of the particle. By definition it is a relativistically invariant quantity.

We now take dt outside the parentheses in the expression for action. Then, by definition, in the parentheses we have the Lagrangian:

$$S = \int \left[-mc^2 \left(1 - \frac{v^2}{c^2} \right)^{1/2} + \frac{e}{c} \mathbf{A}\mathbf{v} - e\varphi \right] dt = \int L \, dt \tag{14.23}$$

From it we obtain, according to the conventional rules, the expressions for momentum and energy. For momentum we have

$$\mathbf{p} = \frac{\partial L}{\partial \mathbf{v}} = \frac{m\mathbf{v}}{(1-v^2/c^2)^{1/2}} + \frac{e}{c}\,\mathbf{A} \equiv \mathbf{p}_0 + \frac{e}{c}\,\mathbf{A} \qquad (14.24)$$

Here, $\mathbf{p}_0$ denotes momentum in the absence of a field.

From (4.1), the energy is

$$E = \mathbf{v}\,\frac{\partial L}{\partial \mathbf{v}} - L = \mathbf{v}\mathbf{p}_0 + \frac{e}{c}\,\mathbf{v}\mathbf{A} + mc^2\,(1-v^2/c^2)^{1/2} - \frac{e}{c}\,\mathbf{v}\mathbf{A} + e\varphi$$

$$= E_0 + e\varphi \qquad (14.25)$$

where E_0 is the energy in the absence of an external field; according to (14.7), it is equal to

$$E_0 = \mathbf{v}\mathbf{p}_0 + mc^2\,(1-v^2/c^2)^{1/2} = \frac{mc^2}{(1-v^2/c^2)^{1/2}}$$

Thus, the term linear with respect to the velocity does not appear in the energy expressed in terms of the velocity. It will be seen here that the Lagrangian is not of the form $T - U$ because it involves a linear term.

From (14.24) and (14.25) we obtain

$$\mathbf{p}_0 = \mathbf{p} - \frac{e}{c}\,\mathbf{A}, \qquad E_0 = E - e\varphi$$

But we already know the expression for E_0 in terms of $\mathbf{p}_0$ from Eq. (14.12), which relates to the energy and momentum of a free particle. Substituting these quantities, expressed in terms of $\mathbf{p}$ and E, into it, we obtain the Hamiltonian of a charge in an external electromagnetic field:

$$\mathscr{H} = \left[m^2c^4 + c^2\left(\mathbf{p} - \frac{e}{c}\,\mathbf{A}\right)^2 \right]^{1/2} + e\varphi \qquad (14.26)$$

In this case we can also write by analogy with the four-dimensional notation (14.19):

$$\left(p_i - \frac{e}{c}\,A_i\right)\left(p_i - \frac{e}{c}\,A_i\right) \equiv \left(p_i - \frac{e}{c}\,A_i\right)^2 = -m^2c^2 \qquad (14.27)$$

Equations of Motion of a Charge in an External Field. Knowing the Lagrangian from (14.23), we write the Lagrange equations for a particle moving in an external electromagnetic field. As always, in the most general form they must be

$$\frac{d}{dt}\,\frac{\partial L}{\partial \mathbf{v}} - \frac{\partial L}{\partial \mathbf{r}} = 0$$

where one vector equation replaces three equations expressed in terms of the components.

The derivative $\partial L/\partial \mathbf{v}$ is equal to $\mathbf{p} = \mathbf{p}_0 + (e/c)\mathbf{A}$, so that the total time derivative of $\mathbf{p}$ is

$$\frac{d}{dt} \frac{\partial L}{\partial \mathbf{v}} = \frac{d\mathbf{p}}{dt} = \frac{d\mathbf{p}_0}{dt} + \frac{e}{c} \frac{d\mathbf{A}}{dt}$$

In order to expand the expression $d\mathbf{A}/dt$, we first write it for one component:

$$\frac{dA_x}{dt} = \frac{\partial A_x}{\partial t} + \frac{\partial A_x}{\partial x} \frac{dx}{dt} + \frac{\partial A_x}{\partial y} \frac{dy}{dt} + \frac{\partial A_x}{\partial z} \frac{dz}{dt} = \frac{\partial A_x}{\partial t} + (\mathbf{v} \cdot \nabla) A_x$$

From this, using the notation $(\mathbf{v} \cdot \nabla)$ (cf. (11.31)) for all three components of $\mathbf{A}$, we rewrite $d\mathbf{p}/dt$ as follows:

$$\frac{d\mathbf{p}}{dt} = \frac{d\mathbf{p}_0}{dt} + \frac{e}{c} \left(\frac{\partial \mathbf{A}}{\partial t} + (\mathbf{v} \cdot \nabla) \mathbf{A} \right)$$

We now compute the derivative $\partial L/\partial \mathbf{r}$, or, what is the same, grad L:

$$\text{grad } L = \frac{e}{c} \nabla(\mathbf{A} \cdot \mathbf{v}) - e \text{ grad } \varphi$$

The gradient $\nabla(\mathbf{A} \cdot \mathbf{v})$ denotes differentiation with respect to coordinates, on which only $\mathbf{A}$ depends explicitly, but not $\mathbf{v}$. Therefore, applying Eq. (11.32), we reduce grad L to the form

$$\frac{\partial L}{\partial \mathbf{r}} \equiv \text{grad } L = \frac{e}{c} (\mathbf{v} \cdot \nabla)\mathbf{A} + \frac{e}{c} (\mathbf{v} \times \text{curl } \mathbf{A}) - e \text{ grad } \varphi$$

Substituting $d\mathbf{p}/dt$ and grad L into the general Lagrange equation and retaining only $d\mathbf{p}_0/dt$ in the left-hand side, we obtain the required equation of motion of a charged particle in an electromagnetic field:

$$\frac{d\mathbf{p}_0}{dt} = \frac{d}{dt} \frac{m\mathbf{v}}{(1 - v^2/c^2)^{1/2}}$$

$$= e \left(-\text{grad } \varphi - \frac{1}{c} \frac{\partial \mathbf{A}}{\partial t} \right) + \frac{e}{c} \mathbf{v} \times \text{curl } \mathbf{A} \qquad (14.28)$$

Recalling Eqs. (12.34) and (12.35), which relate electromagnetic fields and potentials, we rewrite (14.28) in final form:

$$\frac{d}{dt} \frac{m\mathbf{v}}{(1 - v^2/c^2)^{1/2}} = e\mathbf{E} + \frac{e}{c} \mathbf{v} \times \mathbf{H} \qquad (14.29)$$

Equation (14.29) should be seen as the physical definition of an electromagnetic field according to its action on a charged particle. In accordance with the requirement of gauge invariance, the scalar and vector potentials taken by themselves are not involved in equations expressing physical quantities.

It would be extremely difficult, however, to get along without electromagnetic potentials altogether; notably, it would be impos-

sible to introduce the Lagrangian. We shall show that it satisfies the condition imposed on the potentials. Indeed, if the potentials are subjected to the gauge transformation (12.36) and (12.37), that is, if the substitution

$$\mathbf{A} = \mathbf{A}' + \nabla f, \qquad \varphi = \varphi' - \frac{1}{c}\frac{\partial f}{\partial t}$$

is made, we obtain under the integral in (14.23) the term

$$e\,\frac{v}{c}\,\nabla f + e\,\frac{1}{c}\,\frac{\partial f}{\partial t} = \frac{e}{c}\,\frac{df}{dt}$$

But the total derivative of a function of the coordinates and time can always be omitted from a Lagrangian. Accordingly, the equations of motion involve fields, but not potentials.

The vector in the right-hand side of (14.29) is called the *Lorentz force*. It involves, besides the conventional electric force $e\mathbf{E}$ known from electrostatics, the term $(e/c)\mathbf{v} \times \mathbf{H}$, somewhat resembling the Coriolis force. This term is due to the part of the Lagrangian that is linear with respect to velocity.

The magnetic part of the Lorentz force, $(e/c)\mathbf{v} \times \mathbf{H}$, is very like the expression for a force acting on a current in an external magnetic field and can be derived from it in much the same way as, in Section 12, we obtained Maxwell's equations from the fundamental laws of the theory of electromagnetism. However, using this method it is much more difficult to discern the relativistic invariance of the expressions.

The Equations of Motion of a Charge in Four-Dimensional Form. We shall now show that in developing Eq. (14.29) from the invariance principle (14.22) we did not lose the relativistic invariance of the result, even though it would appear that time was factored out in writing (14.23) and is involved in the relationship nonsymmetrically with the coordinates. In particular, unlike the coordinates, it does not vary.

We start with the definition of an electromagnetic field in four-dimensional notation. The components of a magnetic field $\mathbf{H}$ are connected with the vector potential components in the following way:

$$H_x = \frac{\partial A_y}{\partial z} - \frac{\partial A_z}{\partial y}\,, \qquad H_y = \frac{\partial A_x}{\partial z} - \frac{\partial A_z}{\partial x}\,, \qquad H_z = \frac{\partial A_y}{\partial x} - \frac{\partial A_x}{\partial y}$$

Hence, with the help of the components of the four-vector, A_i, involved in (14.22), the magnetic field can be represented as follows:

$$H_x = \frac{\partial A_3}{\partial x_2} - \frac{\partial A_2}{\partial x_3}\,, \qquad H_y = \frac{\partial A_1}{\partial x_3} - \frac{\partial A_3}{\partial x_1}\,, \qquad H_z = \frac{\partial A_2}{\partial x_1} - \frac{\partial A_1}{\partial x_2}$$

We now introduce a four-dimensional antisymmetric tensor F_{ik}, connected with the vector and scalar potentials (or with the four-vector potential) by the equations

$$F_{ik} = \frac{\partial A_k}{\partial x_i} - \frac{\partial A_i}{\partial x_k} \qquad (14.30)$$

Substituting $A_4 = i\varphi$, and $x_4 = ict$ into this, we obtain the components of tensor F_{ik} involving 4 as one of the indices:

$$F_{14} = \frac{\partial A_4}{\partial x_1} - \frac{\partial A_1}{\partial x_4} = i\,\frac{\partial \varphi}{\partial x} - \frac{1}{ic}\,\frac{\partial A_x}{\partial t}$$

$$= \frac{1}{i}\left(-\frac{\partial \varphi}{\partial x} - \frac{1}{c}\,\frac{\partial A_x}{dt}\right) = \frac{E_x}{i}$$

$$F_{24} = \frac{E_y}{i}, \qquad F_{34} = \frac{E_z}{i}$$

We finally write the whole tensor in matrix form:

$$\begin{pmatrix} 0 & H_z & -H_y & -iE_x \\ -H_z & 0 & H_x & -iE_y \\ H_y & -H_x & 0 & -iE_z \\ iE_x & iE_y & iE_z & 0 \end{pmatrix} = \begin{pmatrix} F_{11} & F_{12} & F_{13} & F_{14} \\ F_{21} & F_{22} & F_{23} & F_{24} \\ F_{31} & F_{32} & F_{33} & F_{34} \\ F_{41} & F_{42} & F_{43} & F_{44} \end{pmatrix} \qquad (14.31)$$

The equality of the matrices signifies term-by-term equality of the components. Thus, in four-dimensional notation electric and magnetic fields constitute a single antisymmetric tensor. By definition, it has zeros along its principal diagonal, while the components symmetrically located with respect to the principal diagonal have opposite signs.

Now we multiply the first equation in (14.29) by dt/ds and expand its right-hand side in the components of the magnetic field:

$$\frac{dp_x}{ds} = eE_x\,\frac{dt}{ds} + \frac{e}{c}\,\frac{dy}{ds}\,H_z - \frac{e}{c}\,\frac{dz}{ds}\,H_y$$

Substituting p_x from (14.14), and the electromagnetic field components from (14.31), we arrive at the equation

$$mc\,\frac{d^2 x_1}{ds^2} = \frac{e}{c}\,F_{1k}\,\frac{dx_k}{ds}$$

The notation is similar for the other two equations.

Let us combine the obtained set in one four-dimensional equation. What is the meaning of the fourth equation? Since there were only three relationships in (14.29), we must show that the fourth is a corollary of them.

For this, multiply (14.29) by **v**. In the left-hand side we obtain

$$\mathbf{v}\,\frac{d\mathbf{p_0}}{dt} = \frac{\partial E_0}{\partial \mathbf{p_0}}\,\frac{d\mathbf{p_0}}{dt} = \frac{dE_0}{dt}$$

In the right-hand side the product of the velocity times the component of the Lorentz force vanishes, because $\mathbf{v}\times\mathbf{H}$ and $\mathbf{v}$ are mutually perpendicular. There remains the equation

$$\frac{dE_0}{dt} = e\,(\mathbf{E}\cdot\mathbf{v}) \tag{14.32}$$

Substituting $E_0 = (mc^2/i)(dx_4/ds)$ and the electric field components from (14.31) into it, we observe (after multiplying by dt/ds) that we really obtain the fourth equation for a four-dimensional set:

$$mc\,\frac{d^2x_i}{ds^2} = \frac{e}{c}\,F_{ik}\,\frac{dx_k}{ds} \tag{14.33}$$

In the left-hand side of Eq. (14.32) we have the change in the particle's energy in unit time, that is, the work done on it in that time. On the right-hand side we have the scalar product of the force acting on the particle multiplied by its velocity: the conventional expression of work per unit time. This quantity is known as the *Lorentz work*.

We have thus obtained a relativistically invariant formulation of the equations of mechanics of a charged particle in an electromagnetic field. The essential physical distinction from the equations of Newtonian mechanics is that in the case of an electromagnetic field we cannot introduce the concept of the interaction energy of the particles in the system. Each particle interacts directly only with the electromagnetic field, according to the fact that in relativity theory only short-range action is possible.

Equations (14.29) or (14.33) can be solved only if the external electromagnetic field acting on the particle is known. To have a complete set of electrodynamic equations we must learn, in turn, to find the field from the given motion of the charges. In principle, the problem is solved with the help of Maxwell's equations obtained in Section 12. A relativistically invariant derivation will be presented in the next section. In Section 20 we shall point out certain difficulties due to the application of the simultaneous set of Maxwell's equations and (14.33).

EXERCISES

1. A fast proton possessing energy E collides with a stationary proton. Determine the portion of the energy of the colliding proton that can be dissipated in an inelastic process (for example, the creation of a proton-antiproton pair).

Solution. The energy dissipation of the colliding particles in an inelastic process is greatest in the reference frame in which the total momentum of both particles is zero (see Sec. 6). Let the momentum of one particle in this system be p_0, of the other $-p_0$, and the energy of each $E_0 = (m^2c^4 + c^2p_0^2)^{1/2}$. We proceed from the invariance of the four-dimensional product $p_k^{'(1)} p_k^{(2)} = p_{0k}^{(1)} p_{0k}^{(2)}$. Since in the laboratory reference frame the momentum of the stationary proton is zero while its energy is mc^2, we obtain

$$mc^2 E = E_0^2 + c^2 p_0^2$$

From this it follows that, in a centre-of-mass reference frame the required total energy is equal to

$$\frac{1}{E} [2mc^2 (E + mc^2)]^{1/2}$$

Thus, the relative proportion of the energy of the colliding proton that may be dissipated in an inelastic process is inversely proportional to the square root of its initial energy. That is why at present such attention is being given to the development of accelerators with colliding beams, in which the laboratory reference system and centre-of-mass reference coincide.

2. Consider the collision of a particle of zero mass with a stationary particle of mass m. Determine the energy of the incident particle after the collision, if its energy prior to the collision, E, and the deflection angle, ϑ, are known.

Answer.

$$E' = \frac{E}{1 + (E/mc^2)(1 - \cos \vartheta)}$$

3. Find the dependence of the velocity of a rocket upon the mass of the burned and ejected fuel, if its initial mass M_0 and the velocity v of the ejected particles of the fuel relative to the rocket are known.

Solution. Let the sought velocity of the rocket be V, its initial mass, M_0, and its instantaneous mass, M. If a certain quantity dm of the fuel is ejected, then, relative to a fixed reference frame, the momentum conservation law is written as follows:

$$d \frac{MV}{(1 - V^2/c^2)^{1/2}} = \frac{v' \, dm}{(1 - v'^2/c^2)^{1/2}}$$

Here the velocity v' of the ejected fuel relative to the fixed reference frame is

$$v' = \frac{v - V}{1 - vV/c^2}$$

This takes account of the fact that v' and V are oppositely directed. Substituting v' into the momentum-conservation equation, we obtain

$$d \frac{MV}{(1 - V^2/c^2)^{1/2}} = \frac{(v - V) \, dm}{(1 - v^2/c^2)^{1/2} (1 - V^2/c^2)^{1/2}}$$

The energy-conservation law is more conveniently written in a reference frame fixed with respect to the rocket:

$$dMc^2 = -\frac{c^2\,dm}{(1-v^2/c^2)^{1/2}}$$

After eliminating dm and carrying out the necessary cancellations, we have

$$\frac{dM}{M} = -\frac{dV}{v\,(1-V^2/c^2)}$$

whence, after integrating, we arrive at the required equation:

$$\frac{M}{M_0} = \frac{(1-V/c)^{c/2v}}{(1+V/c)^{c/2v}}$$

If total annihilation of matter occurs in the rocket, and the ejected particles are photons (light quanta), then

$$\frac{M}{M_0} = \left(\frac{1-V/c}{1+V/c}\right)^{1/2}$$

In the nonrelativistic limit, when $v \ll c$,

$$\frac{M}{M_0} = e^{-V/v}$$

4. A stationary particle of mass m decays into two particles of mass m_1 and m_2. Determine the energies of the end products.

Answer.

$$E_1 = \frac{m_1^2 - m_2^2 + m^2}{2m}, \qquad E_2 = \frac{m_2^2 - m_1^2 + m^2}{2m}$$

5. Develop the equations of motion of a charge in an electromagnetic field directly from the invariant expression for action (14.22).

Solution. The variation of action has the form

$$\delta S = \int \left(-mc\delta\,ds + \frac{e}{c}\frac{\partial A_k}{\partial x_i}\,\delta x_i\,dx_k + \frac{e}{c}\,A_k\,d\delta x_k\right) = 0$$

Taking advantage of the fact that $ds^2 = -dx_i^2$, we find the variation of the interval

$$\delta\,ds = -\frac{dx_i\,d\delta x_i}{ds}$$

Transforming the variation differentials by parts, we obtain

$$\delta S = \int \left(-mc\,d\frac{dx_i}{ds} + \frac{e}{c}\frac{\partial A_k}{\partial x_i}\,dx_k - \frac{e}{c}\frac{\partial A_i}{\partial x_k}\,dx_k\right)\delta x_i$$

Equating the factors of the variations to zero, we arrive at the set of equations (14.33):

$$mc\frac{d^2x_i}{ds^2} = \frac{e}{c}\left(\frac{\partial A_k}{\partial x_i} - \frac{\partial A_i}{\partial x_k}\right)\frac{dx_k}{ds} = \frac{e}{c}\,F_{ik}\frac{dx_k}{ds}$$

6. Find the scalar and vector potentials of a freely moving charge.

Solution. In its reference frame, the charge creates only an electro-static field, the scalar potential of which is equal to e/r_0. There is no magnetic field in this frame, so that the vector potential is equal to zero. Transforming the scalar potential as the fourth component of the vector according to the general equations (13.33) and (13.34), and taking into account that A_x in the right-hand sides of these equations vanishes, we obtain

$$\varphi = \frac{\varphi_0}{(1 - v^2/c^2)^{1/2}} = \frac{e}{r_0 \, (1 - v^2/c^2)^{1/2}}$$

$$\mathbf{A} = \frac{\varphi_0 \mathbf{v}/c}{(1 - v^2/c^2)^{1/2}} = \frac{e\mathbf{v}}{r_0 c \, (1 - v^2/c^2)^{1/2}}$$

Furthermore, we must express r_0 in terms of the coordinates in the system relative to which the charge is moving:

$$r_0 = (x_0^2 + y_0^2 + z_0^2)^{1/2} = \left(\frac{(x - vt)^2}{1 - v^2/c^2} + y^2 + z^2 \right)^{1/2}$$

Instead of vt we may substitute ξ, that is, the abscissa of the moving charge. The electromagnetic disturbance arrives at time t at the point with coordinates x, y, z, not from point $(\xi, 0, 0)$, where it is located at the given instant, but from point $(\xi', 0, 0)$, where it was located at the time of the emission of the disturbance. If the distance from point $(\xi', 0, 0)$ to the point with coordinates x, y, z is R', the time required for the electro-magnetic disturbance to travel along it is R'/c. The charge moving with the velocity v would have taken the same time to travel the path $(\xi - \xi')/v$. From this we have the equation

$$\frac{\xi - \xi'}{v} = \frac{[(x - \xi')^2 + y^2 + z^2]^{1/2}}{c} = \frac{R}{c}$$

Here, $\xi = vt$, that is, the charge's abscissa at the current time t. The differ-ence $x - \xi'$ is the projection of vector $\mathbf{R}'$ on the direction of the velocity, $(\mathbf{v} \cdot \mathbf{R}')/v$. Substituting the obtained expressions into r_0, we finally obtain

$$\varphi = \frac{e}{R' - \mathbf{v}\mathbf{R}'/c} \, , \qquad \mathbf{A} = \frac{e\mathbf{v}}{c \, (R' - \mathbf{v}\mathbf{R}'/c)}$$

7. Find the motion of a charge in a constant uniform magnetic field.

Solution. If the field is in the direction of the z axis, the equations of motion are of the following form:

$$\frac{dp_x}{dt} = \frac{e}{c} \, \frac{dy}{dt} \, |\, \mathbf{H}\,|, \qquad \frac{dp_y}{dt} = -\frac{e}{c} \, \frac{dx}{dt} \, |\, \mathbf{H}\,|, \qquad \frac{dp_z}{dt} = 0$$

Since the magnetic field does not do work on the charge, $p^2 = $ constant, $p_z = $ constant, $p_x^2 + p_y^2 = $ constant, and

$$p_x = \frac{mv_x}{(1 - v^2/c^2)^{1/2}} = \frac{Ev_x}{c^2}$$

We look for the coordinates x and y in the form:

$$x = r \cos \omega t, \qquad y = r \sin \omega t$$

For r and ω the following expressions result:

$$r = \frac{Ev}{ec \,|\, \mathbf{H} \,|} \,, \qquad \omega = \frac{ec \,|\, \mathbf{H} \,|}{E}$$

The particle moves along a helix. For small velocities, ω reduces to the constant value $e|\mathbf{H}|/mc$.

8. Find the motion of a charge in a constant uniform electric field.
Solution. The equations of motion are

$$\frac{dp_x}{dt} = e \,|\, \mathbf{E} \,|, \qquad \frac{dp_y}{dt} = 0, \qquad \frac{dp_z}{dt} = 0, \qquad \frac{dE_0}{dt} = e \,|\, \mathbf{E} \,| \frac{dx}{dt}$$

From the last equation we obtain

$$[m^2 c^4 + c^2 \,(p_x^2 + p_y^2 + p_z^2)]^{1/2} - [m^2 c^4 + c^2 \,(p_{x_0}^2 + p_{y_0}^2 + p_{z_0}^2)]^{1/2} = e \,|\, \mathbf{E} \,|\, x$$

From the first equation

$$p_x - p_{x_0} = e \,|\, \mathbf{E} \,|\, t, \qquad p_y - p_{y_0} = 0, \qquad p_z - p_{z_0} = 0$$

These equations together give x as a function of t.

If $p_{z_0} = 0$, then dividing p_x by p_y we have an expression for dx/dy in terms of x (by eliminating t from the energy integral). The trajectory is of the form of a catenary.

9. Express the energy of motion of a charge in an attracting Coulomb field in terms of the adiabatic invariants (action variables).
Solution. Since the potential has only a scalar component, we find from (14.27), after passing to plane motion in polar coordinates:

$$\frac{1}{c^2} \,(E - e\varphi)^2 = m^2 c^2 + p_r^2 + \frac{p_\varphi^2}{r^2}, \quad J_\varphi = p_\varphi$$

From this the radial action variable is

$$J_r = \frac{1}{2\pi} \int p_r \, dr = \frac{1}{2\pi} \int \left[-\left(m^2 c^2 - \frac{E^2}{c^2} \right) + \frac{2EZe^2}{rc} \right.$$
$$\left. - \frac{1}{r^2} \left(J_\varphi^2 - \frac{Z^2 e^4}{c^2} \right) \right]^{1/2} dr$$

The integral is taken over the whole domain in which the radicand is real, taking into account that $E < mc^2$, since otherwise the motion is infinite. The assumption $E < mc^2$ corresponds to $E < 0$ in Newtonian mechanics, that is, to finite motion.

Integrating, we find

$$J_r = \frac{Ze^2 E}{c \,(m^2 c^4 - E^2)^{1/2}} - \left(J_\varphi^2 - \frac{Z^2 e^4}{c^2} \right)^{1/2}$$

whence we express the energy in terms of the action variables J_r and J_φ:

$$E = mc^2 \left[1 + \frac{Z^2 e^4}{c^2 \left[J_r + (J_\varphi^2 - Z^2 e^4/c^2)^{1/2} \right]^2} \right]^{-1/2}$$

The derivatives $\partial E/\partial J_r$ and $\partial E/\partial J_\varphi$ are not equal, so that the path of the charge is not closed. It has the shape of an ellipse whose axes are rotating (a rosette).

15

ACTION OF AN ELECTROMAGNETIC FIELD

The Lorentz Transformations of Field Components. In this section it will be shown that Maxwell's equations can be treated as mechanical equations of motion applied to electromagnetic fields. For this they must be developed from the general principle of mechanics, just as the equations of the mechanics of mass points were developed in Section 2 from the principle of least action.

To formulate the corresponding principle for electromagnetic fields we must proceed from similar requirements: invariance of the equations with respect to spatial and temporal displacements of the coordinate system, its rotations, and transformations to other inertial frames of reference.

The requirement of invariance with respect to spatial and temporal displacements reduces to the requirement that the action of the field does not involve explicit functions of the coordinates or time. In other words, it can depend only on quantities describing the field, just as the action of a closed mechanical system depends only on its generalized coordinates and velocities.

As for the condition of invariance with respect to spatial rotations and transformations to other inertial reference frames, in Section 13 it was shown that both requirements are best satisfied by writing the equations in four-dimensional tensor notation.

We should therefore begin with the question of transforming the electromagnetic field components so as to see what invariant quantities can be obtained.

As was shown in Section 14, a field is a four-dimensional antisymmetric tensor of rank 2 the components of which are given by Eq. (14.31). Let us apply to it the ordinary Lorentz transformation for the case when the relative velocity of the reference frames is directed along the x axis. Such a transformation affects only the first and fourth tensor indices according to Eqs. (13.33) and (13.34).

It cán be seen at once from this that the magnetic field component along the relative velocity, $H_x = F_{23}$, does not transform at all: it has the tensor indices 2 and 3.

Take the magnetic field component H_z and the electric field component E_y. In tensor notation these are F_{12} and F_{42}/i. The index *2* is not affected by the transformation, while indices *4* and *1* are transformed according to the general rules (13.33) and (13.34). Thus we obtain the transformation equations of all four lateral electromagnetic field components:

$$H'_z = \frac{H_z - V E_y/c}{(1 - V^2/c^2)^{1/2}} \tag{15.1a}$$

$$H'_y = \frac{H_y + V E_z/c}{(1 - V^2/c^2)^{1/2}} \tag{15.1b}$$

$$E'_y = \frac{E_y - V H_z/c}{(1 - V^2/c^2)^{1/2}} \tag{15.2a}$$

$$E'_z = \frac{E_z + V H_y/c}{(1 - V^2/c^2)^{1/2}} \tag{15.2b}$$

There remains the longitudinal electric field component $E_x = iF_{14}$, for which both indices *1* and *4* transform. In this case the transformation can be considered as a rotation in two-dimensional space with coordinates x_1 and x_4. Then F_{14} is an antisymmetric tensor of rank 2 in two-dimensional space. But it was shown in Section 11 that a completely antisymmetric tensor whose rank is equal to the number of spatial dimensions is invariant under rotations. What was said of three-dimensional space can be literally transferred to any number of dimensions, including two. Hence, F_{14} is invariant under rotations of the form (13.33) and (13.34). Of course, this can also be demonstrated with the help of computations according to the general formulas of tensor transformations.

Thus, for both longitudinal field components we have

$$H'_x = H_x \tag{15.1c}$$

$$E^{\cdot}_x = E_x \tag{15.2c}$$

Let us find the quantities invariant with respect to the transformations. An arbitrary tensor A_{ik} has an invariant A_{ii}, that is, the sum of the diagonal components. This quantity, which possesses only dummy indices and is called the *trace of a tensor*, is invariant. But since tensor F_{ik} is antisymmetric, all its diagonal elements vanish (cf. (14.31)). Consequently $F_{ii} = 0$.

Thus, there is no linear invariant; but apparently there must be a quadratic invariant $F_{ik}F_{ik}$, that is, the sum of the squares of all

the tensor components:

$$F_{ik}F_{ik} = F_{ik}^2 = 2\,(\,|\,\mathbf{H}\,|^2 - |\,\mathbf{E}\,|^2) \tag{15.3}$$

This quantity is of fundamental importance for the subsequent treatment.

With the help of a completely antisymmetric tensor of rank 4 we can construct one more invariant quantity:

$$\varepsilon_{iklm}F_{ik}F_{lm} \tag{15.4}$$

Since ε_{iklm} is an invariant tensor, we have, of course, obtained a relativistically invariant quantity. Let us expand it with the help of the electromagnetic field components:

$$\varepsilon_{iklm}F_{ik}F_{lm} = 8i\,(E_xH_x + E_yH_y + E_zH_z) = 8i\,(\mathbf{E}\cdot\mathbf{H})$$

Tensor ε_{iklm} has $4! = 24$ components.

We have thus found one more quantity which does not change under the Lorentz transformations—the scalar product $(\mathbf{E}\cdot\mathbf{H})$.

Scalars and Pseudoscalars. Vectors and Pseudovectors. We shall now show that the scalar product $(\mathbf{E}\cdot\mathbf{H})$ is in a sense not a true scalar. Namely, $(\mathbf{E}\cdot\mathbf{H})$ reverses its sign if the signs of all the coordinates are reversed: $x' = -x$, $y' = -y$, $z' = -z$.

This transformation is known as *inversion of the coordinate system*, or *mirror reflection*. Indeed, it transforms a right-handed system into a left-handed system; but that is just how a right-hand system appears in a mirror (the right hand in a mirror appears as the left hand).

It is not hard to see that no rotation of the coordinate axes can lead to an inversion. The transformation matrix corresponding to an inversion must be written in the form

$$\begin{pmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix}$$

The determinant of this matrix is equal to -1, whereas the corresponding determinant for a rotation transformation is equal to $+1$. There exists no continuous transformation from $+1$ to -1.

Vectors behave in different ways with respect to inversions. The velocity vector $\mathbf{v} = d\mathbf{r}/dt$ changes sign together with $\mathbf{r}$. The momentum vector $\mathbf{p} = m\mathbf{v}$ apparently also changes its sign. The force vector $\mathbf{F} = \dot{\mathbf{p}}$ possesses the same property. All these vectors are known as *true*, or *polar*, *vectors* or simply vectors.

The angular momentum vector $\mathbf{M} = \mathbf{r} \times \mathbf{p}$, the components of which involve the products of the components of $\mathbf{r}$ and $\mathbf{p}$, does not apparently change sign; nor does the torque vector $\mathbf{K} = \mathbf{r} \times \mathbf{F}$.

Such vectors are called *axial* or *pseudovectors*: they behave like vectors in rotations of the coordinate system, but differently in inversions. It was shown in Section 11 that **M** and **K**, that is, actually vector products of true vectors, should be defined as antisymmetric tensors of rank 2. They possess vector properties only in three-dimensional space, because the number of components of an antisymmetric tensor and a vector coincide. Inversion reveals the nonvector nature of **M** and **K**.

It is not hard to see that angular velocity is a pseudovector. It is related to angular momentum by the equality

$$M_\alpha = I_{\alpha\beta}\omega_\beta$$

(see Sec. 9). But the components of the inertia tensor depend on the coordinates quadratically and do not change sign in inversions, so that in this respect angular velocity behaves like angular momentum. This is because the direction of the angular velocity vector was chosen arbitrarily (Sec. 8).

We shall now show that the vector potential **A** is a true vector. It appears in the expression for action (14.23) in a scalar product with the velocity vector **v**, a true vector. From experience we know that the equations of mechanics are invariant under inversion: their form does not change in the substitution of a left-handed system for a right-handed system. Consequently, the scalar product $(\mathbf{A} \cdot \mathbf{v})$ involved in the Lagrangian should not change sign under inversion. For that **A** must be a true vector. Since (14.23) can be regarded as a definition of **A**, the vector properties of **A** follow precisely from it.

But it is then obvious that the magnetic field **H** is a pseudovector, because the del, $\nabla = \partial/\partial r$, is obviously a true vector, which transforms under inversion like **r**, while the magnetic field

$$\mathbf{H} = \operatorname{curl} \mathbf{A} \equiv \nabla \times \mathbf{A}$$

An electric field, defined as $\mathbf{E} = -\nabla\varphi - (1/c)(\partial\mathbf{A}/\partial t)$, is a true vector by virtue of the fact that ∇ and **A** are true vectors. The scalar product $(\mathbf{E} \cdot \mathbf{H})$ of vector **E** multiplied by pseudovector **H** changes sign under inversion, and is thus a pseudoscalar; $(\mathbf{E} \cdot \mathbf{H})$ remains invariant only in rotations of the coordinate system, but not in inversions.

The four-dimensional tensor definition of $(\mathbf{E} \cdot \mathbf{H})$ also indicates its pseudoscalar nature: each term includes three spatial indices and one temporal, *4*, so that a change in the sign of the three coordinates changes the sign of $(\mathbf{E} \cdot \mathbf{H})$.

One might think that the equations of electrodynamics should not be invariant in inversions: since in one case the right-hand rule is applied, while in the other it is the left-hand rule. But actually these rules derive from the convention of defining the signs—or poles—of a magnetic field. The laws of electrodynamics would

remain essentially unaffected if we changed their names and at the same time interchanged the respective hand rules.

The law of right and left symmetry is not universal. There is in nature a special class of interactions of a nonelectromagnetic nature—the weak interactions—in which there is no inversion symmetry. This symmetry is an expression of a specific property of a specific type of interaction, which can be established only experimentally.

We have thus restricted ourselves to two invariant quantities: $|H|^2 - |E|^2$ and $(E \cdot H)$. All the others can be expressed in terms of them.

The Field Linearity of Maxwell's Equations. A true invariant can be formed from $(E \cdot H)$ by squaring it. Of course, it is never apparent in advance why such a quantity (or the square of the invariant $|H|^2 - |E|^2$ either) cannot enter the expression for the action of an electromagnetic field, as well as terms of higher powers. However, if terms higher than quadratic with respect to the field were included, the equations of electrodynamics obtained by action variation would involve nonlinear terms, that is, the squares of fields, their products, etc.

The basic difference between a nonlinear and linear equation is that the sum of the two solutions of a nonlinear equation is not a solution, because cross terms of both solutions appear. But it is well known that two electromagnetic waves propagating in a vacuum are simply added, without distorting each other. In nonlinear theory, the velocity of a wave depends on its amplitude, whereas in electrodynamics all disturbances propagate in vacuum with the same speed.

Proceeding from this experimental fact, we must select only the quadratic invariant of the electromagnetic field for the field equations to be linear. With two quadratic invariants we could have taken their linear combination in the action expression. But $(E \cdot H)$ being a pseudoscalar, the combination would alter the relative sign between it and $|H|^2 - |E|^2$. If, therefore, we retain $(E \cdot H)$, the equations of electrodynamics will be of different form in right-handed and left-handed coordinate systems, and no redefinition of the sign of the magnetic field can help.

Consequently, there remains only one quadratic quantity, $|H|^2 - |E|^2$, which can be involved in the expression for the action of an electromagnetic field.

Another quadratic quantity, $A_i A_i = |A|^2 - \varphi^2$, can be obtained from the vector potential, but it is not guage invariant and also cannot appear in the action formula.

Field-Charge Interaction. We obtained the corresponding expression (14.22) in Section 14. For a separate charge it has the form

$S_{\text{int}} = \int (e/c) A_k \, dx_k$. It is more convenient to go over from a point charge to a spatially distributed one (as shown in Section 12, this can always be done).

To trace the relativistic invariance of the equation, let us show that $\mathbf{j}$ and ρ, that is, the current density and the charge density, together comprise one four-vector. The charge differential de is by definition invariant:

$$de = \rho d^{(3)}\tau \tag{15.5}$$

Let the charges be at rest in the reference frame for which Eq. (15.5) was written. We denote this by the subscript 0 for ρ and $d^{(3)}\tau$. Since charge is an invariant quantity, we can write the equation

$$de = \rho_0 d^{(3)}\tau_0 = \rho d^{(3)}\tau \tag{15.6}$$

But as was shown in Exercise 5, Section 13, the volume element $d^{(3)}\tau_0$ moving together with the charged particles contracts relative to a stationary volume element:

$$d^{(3)}\tau = d^{(3)}\tau_0 \, (1 - v^2/c^2)^{1/2} \tag{15.7}$$

It follows from this that the charge density ρ is connected with its density ρ_0 in the stationary reference frame by the relationship

$$\rho = \frac{\rho_0}{(1 - v^2/c^2)^{1/2}} \tag{15.8}$$

The coefficient involved here is expressed in terms of the infinitesimal interval associated with the motion of the charges (see Sec. 14):

$$\frac{1}{(1 - v^2/c^2)^{1/2}} = \frac{c \, dt}{ds}$$

Consequently, the charge density can, in turn, be represented as the fourth component of a four-vector:

$$\rho = \rho_0 \frac{c \, dt}{ds} = \frac{\rho_0}{i} \frac{dx_4}{ds} \equiv \frac{j_4}{ic} \tag{15.9a}$$

Then the current density vector is[4]

$$j_\alpha = c \rho_0 \frac{dx_\alpha}{ds} \tag{15.9b}$$

We now substitute the differentials $\rho_0 \, (dx_k/ds) \, d^{(3)}\tau$ for the expressions $e \, dx_k$ in S_{int} to get

$$S_{\text{int}} = \frac{1}{c} \int j_k A_k \, d^{(3)}\tau \, dt = \frac{1}{c} \int j_k A_k \, d^{(4)}\tau \tag{15.10}$$

---

[4] Note that the Greek indices assume only the values 1, 2, 3.

The Variation Principle for the Electromagnetic Field. The principal task of this section is to show that, like the equations of mechanics, Maxwell's equations are equivalent to a certain variation principle. Although electrodynamics cannot be reduced to a mechanical system of material particles or a continuous medium based on Newton's laws, there exists a far-reaching analogy between mechanics and electrodynamics based on Hamilton's principle. This, it should be noted, is not a question of a formal, superficial analogy. The variation principle makes it possible to define such electromagnetic field quantities as momentum and energy, which are conserved for it, not separately but together with the respective quantities for particles. This is the basis of conservation laws of a more general form, which hold in closed systems comprising charged particles and the field created by them.

In defining the action for a system of mass points, summation over their coordinates is carried out. To use a term from mechanics, an electromagnetic field is a system with an infinite number of degrees of freedom, because to define it fully it is necessary to state the values of all its components at all points of space where the field is not zero. But the points of space constitute an uncountable set, that is, they cannot be numbered in any order. Therefore, in the case of an electromagnetic field, summation is replaced by integration over continuously varying parameters: the coordinates of the points defining the field. The values of the coordinates are analogous to the numbers listing the degrees of freedom of a mechanical system.

The generalized coordinates of a field are given by the values of the vector potential according to the correspondence $q_k(t) \rightarrow \mathbf{A}(\mathbf{r}, t)$. The time derivatives of the vector potential are involved only in the electric field expression. But the time derivatives of the coordinates are involved in the Lagrangian of a mechanical system only through the kinetic energy, that is, with the positive sign. Hence, in substituting the invariant $F_{il}F_{il}$ into the Lagrangian of an electromagnetic field, it must be multiplied by such a factor that the electric field be defined by a positive term.

The numerical value of this coefficient is equal to $-1/(16\pi)$, which corresponds to the Gaussian system of units. Sometimes $-1/4$ is taken, in which case it is said to be Heaviside units of measurement of electromagnetic quantities.

Taking into account the interaction of field and charges, that is, the term (15.10), we can now write the action for an electromagnetic field thus:

$$S = \int \left(-\frac{F_{ik}F_{ik}}{16\pi} + \frac{A_i j_i}{c} \right) d^{(4)}\tau \equiv \int L \, dt \qquad (15.11)$$

Factoring out the time integration in the four-dimensional volume element, we obtain the Lagrangian, which is easily rewritten in

three-dimensional notation:

$$L = \int \left(\frac{|\mathbf{E}|^2 - |\mathbf{H}|^2}{8\pi} + \frac{\mathbf{Aj}}{c} - \rho\varphi \right) d^{(3)}\tau \qquad (15.12)$$

Taking $\partial\mathbf{A}/\partial t$ as the generalized velocity, we can derive the second pair of Maxwell's equations from the Lagrangian in three-dimensional form and then use the matrix (14.31) to determine their relativistic invariance.

Instead, however, we shall perform the variation (15.11) directly in four-dimensional form. The distribution of currents and charges, that is j_k, is assumed to be given, so that only A_k is varied. For the variation of action we obtain

$$\delta S = \int \left[-\frac{F_{ik}}{8\pi} \left(\frac{\partial\delta A_k}{\partial x_i} - \frac{\partial\delta A_i}{\partial x_k} \right) + \frac{j_k}{c}\,\delta A_k \right] d^{(4)}\tau = 0$$

Account was taken in the variation that $F_{ik} = (\partial A_k/\partial x_i) - (\partial A_i/\partial x_k)$. In the second term we redesignate the dummy indices, so that i becomes k, and k becomes i. Then this term differs from the former only in the order of the indices in F_{ki}. But since F_{ki} is an antisymmetric tensor, in interchanging the indices we must change the sign, after which both first terms in the varied expression are reduced. We transform the expression $F_{ik}\,(\partial\delta A_k/\partial x_i)$ by parts to get

$$F_{ik}\frac{\partial\delta A_k}{\partial x_i} = \frac{\partial}{\partial x_i} F_{ik}\delta A_k - \delta A_k \frac{\partial F_{ik}}{\partial x_i}$$

We integrate the derivative $(\partial/\partial x_i)(F_{ik}\delta A_k)$ over the corresponding variable $d^{(4)}\tau$. As always in developing equations of motion, the variation at the integration limits should be put equal to zero. We thus finally reduce the variation δS to the form

$$\delta S = \int \left[-\frac{1}{4\pi}\frac{\partial F_{ki}}{\partial x_i} + \frac{j_k}{c} \right] \delta A_k \, d^{(4)}\tau = 0 \qquad (15.13)$$

For the requirement $\delta S = 0$ to be satisfied the expression in brackets multiplied by the arbitrary variation δA_k should vanish at every point of the four-dimensional volume $d^{(4)}\tau$. From this we obtain the required equations:

$$\frac{\partial F_{ki}}{\partial x_i} = \frac{4\pi}{c} j_k \qquad (15.14)$$

Substituting the components of the four-dimensional tensor F_{ki} from (14.31), we arrive at Maxwell's equations (12.32) and (12.33).

To round out the picture let us present the first pair of equations in four-dimensional form and also give the equations for potential. We write three equations:

$$F_{ik} = \frac{\partial A_k}{\partial x_i} - \frac{\partial A_i}{\partial x_k}, \quad F_{kl} = \frac{\partial A_l}{\partial x_k} - \frac{\partial A_k}{\partial x_l}, \quad F_{li} = \frac{\partial A_i}{\partial x_l} - \frac{\partial A_l}{\partial x_i}$$

differentiate the first with respect to x_l, the second with respect to x_i, and the third with respect to x_k and add them together. Then the right-hand sides cancel out, leaving

$$\frac{\partial F_{ik}}{\partial x_l} + \frac{\partial F_{kl}}{\partial x_i} + \frac{\partial F_{li}}{\partial x_k} = 0 \tag{15.15}$$

It is again not hard to see that these four equations (according to the number of ways of selecting three indices out of four) are equivalent to (12.30) and (12.31).

We write the operation

$$\nabla^2 - \frac{1}{c^2} \frac{\partial^2}{\partial t^2} = \frac{\partial^2}{\partial x_1^2} + \frac{\partial^2}{\partial x_2^2} + \frac{\partial^2}{\partial x_3^2} + \frac{\partial^2}{\partial x_4^2}$$

as $\quad \dfrac{\partial}{\partial x_i} \dfrac{\partial}{\partial x_i} = \dfrac{\partial}{\partial x_i^2}$, sometimes denoted by a square by analogy with ∇. Then instead of (12.43) and (12.44) we obtain for the potential component

$$\square A_k \equiv \frac{\partial^2 A_k}{\partial x_i^2} = -\frac{4\pi}{c} j_k \tag{15.16}$$

Finally, the Lorentz condition (12.42) is written in four-dimensional form as

$$\frac{\partial A_i}{\partial x_i} = 0 \tag{15.17}$$

The Energy-Momentum Tensor of the Electromagnetic Field. We shall now show how Maxwell's equations (15.14) and (15.15), together with the equations of motion of a charged particle in electromagnetic field (14.33), assure satisfaction of the basic mechanical conservation laws: energy, linear momentum, and angular momentum, thereby finally confirming the legitimacy of treating a field as a mechanical system.

After a certain alteration of indices multiply equation (15.14) by F_{li} to get

$$F_{li} \frac{\partial F_{ik}}{\partial x_k} = \frac{4\pi F_{li} j_i}{c}$$

In the obtained equation, redesignate i as k and k as i and add both absolutely equivalent equations. We transform the left-hand sides by parts, after which we obtain

$$\frac{\partial}{\partial x_k} F_{li} F_{ik} + \frac{\partial}{\partial x_i} F_{lk} F_{ki} - F_{ik} \left(\frac{\partial F_{li}}{\partial x_k} + \frac{\partial F_{kl}}{\partial x_i} \right) = \frac{8\pi F_{li} j_i}{c}$$

In combining the two latter terms on the left we made use of the fact that

$$F_{ik} = -F_{ki}, \qquad F_{lk} = -F_{kl}$$

The first two terms on the left can be reduced if in the second term we redesignate the indices in reverse of what we just did. In accordance with the first group of Maxwell's equations (15.15), we replace the quantity in parentheses in the third term by $-\partial F_{ik}/\partial x_l$. Then this term becomes

$$F_{ik}\frac{\partial F_{ik}}{\partial x_l} = \frac{1}{2}\frac{\partial}{\partial x_l}F_{nm}^2 = \frac{1}{2}\,\delta_{kl}\frac{\partial}{\partial x_k}F_{nm}^2$$

Introducing now the notation

$$T_{lk} = \frac{1}{4\pi}\left(F_{li}F_{ki} - \frac{1}{4}\,\delta_{kl}F_{nm}^2\right) \tag{15.18}$$

we write the result of the transformations of the set of Maxwell's equations in the form

$$-\frac{\partial}{\partial x_k}\,T_{lk} = \frac{F_{lk}}{c}\,j_k \tag{15.19}$$

To establish the meaning of this equation, integrate both sides over the three-dimensional volume. First consider the right-hand side. We go over to point charges, for which we must replace the four-dimensional current density vector (see (15.9b)) by its expression in terms of charge density:

$$j_k = c\rho_0\frac{dx_k}{ds}$$

and, besides, replace the charge density in its proper reference system by its density in the stationary system:

$$\rho_0 = \frac{\rho}{c}\frac{ds}{dt}$$

Then integration over the three-dimensional volume is reduced simply to the substitution of $\rho d^{(3)}\tau$ by the total charge e. Hence, the integral of the right-hand side of (15.19) over the three-dimensional volume yields

$$\frac{1}{c}\int F_{lk}j_k d^{(3)}\tau = \frac{e}{c}\,F_{lk}\frac{dx_k}{dt}$$

If we now multiply (14.33) by ds/dt, we find that in the right-hand side of (15.19) we have the total time derivative of the momentum component p_l of the charge or system of charges located in the volume over which the integration was performed.

Consider the left-hand side of Eq. (15.19) after integrating over the volume. It comprises four components corresponding to the dummy index k. In the fourth component the differentiation sign with respect to x_4 is taken outside the integral sign, because dx_4 is a differential of ict, while the integration is over $d^{(3)}\tau$.

As we shall soon see, $T_{\alpha 4}$ also involves an imaginary unit, so that the fourth term of the integrated expression can be written down as the time derivative of the real expression:

$$\frac{\partial}{\partial t} \int \frac{T_{\alpha 4}}{ic} d^{(3)}\tau$$

The remaining three terms have the form of an integral of the divergence over the volume (cf. (11.48)); only in this case the divergence is taken of a tensor rather than a vector. However, it is apparent from the method used in developing the Gauss theorem that it is applicable to the integral of any divergence. Consequently, the first three terms of the integrated equation (15.19) turn into a surface integral. As a result, for the first three values of l, that is, for the spatial components, we have an equation of the form

$$-\left(\frac{\partial}{\partial t} \int \frac{T_{\alpha 4}}{ic} d^{(3)}\tau + \int T_{\alpha\beta}\, dS_\beta\right) = \frac{d}{dt} p_\alpha \qquad (15.20a)$$

while for the fourth, temporal, component we have a similar equation:

$$-\left(\frac{\partial}{\partial t} \int (-T_{44})\, d^{(3)}\tau + \int (-ic T_{4\beta})\, dS_\beta\right) = \frac{dE}{dt} \qquad (15.20b)$$

The equations are written so that all terms should be real, that is not involve an imaginary unit.

In the left-hand side we have the variation of a certain quantity integrated over the volume and added to the flux of a certain other vector quantity across the surface bounding that volume. Together this is equal to the change in the momentum or energy of the charges in unit time. If, for example, the integration surface is so far away that the field on it vanishes, the equations acquire a very simple form

$$\frac{d}{dt}\left(p_\alpha + \int \frac{T_{\alpha 4}}{ic} d^{(3)}\tau\right) = 0 \qquad (15.21a)$$

$$\frac{d}{dt}\left(E + \int (-T_{44})\, d^{(3)}\tau\right) = 0 \qquad (15.21b)$$

that is, each of the quantities under the time derivative sign is constant (the partial derivative multiplying the integral can here be replaced by a total derivative, because in the present case the value of the integral does not depend upon the surface, provided the field vanishes on it).

The system of charges and fields for which Eqs. (15.21a) and (15.21b) were written is closed: since there is no field on its surface, it interacts with nothing (for in electrodynamics there is no such thing as action at distance that could be realized through a surface far away). In closed systems momentum and energy are conserved,

hence we have arrived at the definition of the momentum and energy of an electromagnetic field:

$$p_{\alpha \text{ field}} = \int \frac{T_{\alpha 4}}{ic} \, d^{(3)}\tau \qquad (15.22a)$$

$$E_{\text{field}} = \int (-T_{44}) \, d^{(3)}\tau \qquad (15.22b)$$

The integrands represent the momentum and energy densities, respectively.

Now let us return to the case of the field not being zero at the surface. If we take the surface so that there are no charges within it, Eqs. (15.20a) and (15.20b) take the form

$$\frac{\partial}{\partial t} \int \frac{T_{\alpha 4}}{ic} \, d^{(3)}\tau + \int T_{\alpha \beta} \, dS_{\beta} = 0 \qquad (15.23a)$$

$$\frac{\partial}{\partial t} \int (-T_{44}) \, d^{(3)}\tau + \int \frac{cT_{4\beta}}{i} \, dS_{\beta} = 0 \qquad (15.23b)$$

The first terms in these equations denote the change in the momentum and energy of the field in the given volume. Hence, the second terms denote the momentum and energy fluxes across the surface containing the volume. Equations of similar form are obtained from Eq. (12.18) expressing the charge-conservation law. In integral form it is represented in (12.17).

We have thus revealed the meaning of all the components of the tensor T_{ik}. Let us summarize the results of the foregoing reasoning.

The components $T_{\alpha \beta}$ represent the flux per unit time, of the field momentum component along axis x_{α} across a unit area, the normal to which is directed along axis x_{β}. But momentum transported in unit time is force. Referred to unit area, it yields the normal or tangential stress, depending on whether α and β coincide or not. That is why the spatial tensor $T_{\alpha \beta}$ has a special name, the *Maxwell stress tensor*, which commemorates the fact that it was introduced by Maxwell.

Components $T_{\alpha 4}/(ic)$ denote the density of the spatial momentum component along axis x_{α}.

Components $-icT_{4\alpha}$ represent the energy flux density, that is, the energy transported by the field in unit time across a unit surface the normal of which is directed along axis x_{α}.

Component T_{44} with opposite sign is the energy density of the electromagnetic field.

We shall now express all these components directly in terms of the electromagnetic field components. From Eq. (15.18) we

find

$$T_{11} = \frac{1}{4\pi} \left(F_{1k}F_{1k} - \frac{1}{4} F_{ln}^2 \right)$$

$$= \frac{1}{4\pi} \left(F_{1k}F_{1k} - \frac{1}{2}(|\mathbf{H}|^2 - |\mathbf{E}|^2) \right)$$

$$= \frac{1}{4\pi} \left(H_z^2 + H_y^2 - E_x^2 - \frac{1}{2}(|\mathbf{H}|^2 - |\mathbf{E}|^2) \right)$$

$$= \frac{1}{8\pi} (H_z^2 + H_y^2 - H_x^2 + E_z^2 + E_y^2 - E_x^2) \qquad (15.24a)$$

Similarly, we determine the other components of $T_{\alpha\beta}$ with equal spatial indices. Components with different spatial indices are found as follows:

$$T_{12} = \frac{1}{4\pi}(F_{13}F_{23} + F_{14}F_{24}) = -\frac{1}{4\pi}(E_xE_y + H_xH_y) \quad (15.24b)$$

Components with *4* in the index are equal to

$$T_{41} = T_{14} = \frac{i}{4\pi}(E_yH_z - E_zH_y) = \frac{i}{4\pi}(\mathbf{E} \times \mathbf{H})_x \qquad (15.24c)$$

Hence, the density of the momentum of the field along the x axis is equal to

$$\frac{1}{4\pi c}(\mathbf{E} \times \mathbf{H})_x \qquad (15.25)$$

while the density of the energy flux along the x axis is

$$\frac{c}{4\pi}(\mathbf{E} \times \mathbf{H})_x \qquad (15.26)$$

This vector has a special name, the *Poynting vector*. Finally, the energy density is

$$-T_{44} = -\frac{1}{4\pi} \left(F_{4k}F_{4k} - \frac{1}{2}(|\mathbf{H}|^2 - |\mathbf{E}|^2) \right)$$

$$= \frac{1}{8\pi}(|\mathbf{E}|^2 + |\mathbf{H}|^2) \qquad (15.24d)$$

Let us now write all the tensor components, multiplied by 4π, in the form of an array (they should be seen as arranged four in a row):

$$\frac{1}{2}(E_y^2 + E_z^2 - E_x^2 + H_y^2 + H_z^2 - H_x^2), \quad - \quad (E_xE_y + H_xH_y),$$

$$-(E_xE_z + H_xH_z), \qquad i(\mathbf{E} \times \mathbf{H})_x$$

$$-(E_xE_y + H_xH_y), \qquad \frac{1}{2}(E_x^2 + E_z^2 - E_y^2 + H_x^2 + H_z^2 - H_y^2),$$

$$-(E_yE_z + H_yH_z), \qquad i(\mathbf{E} \times \mathbf{H})_y$$

$$-(E_x E_z + H_x H_z), \qquad -(E_y E_z + H_y H_z),$$

$$\frac{1}{2}(E_x^2 + E_y^2 - E_z^2 + H_x^2 + H_y^2 - H_z^2), \qquad i(\mathbf{E} \times \mathbf{H})_z$$

$$i(\mathbf{E} \times \mathbf{H})_x, \qquad i(\mathbf{E} \times \mathbf{H})_y, \qquad i(\mathbf{E} \times \mathbf{H})_z,$$

$$-\frac{|\mathbf{E}|^2 + |\mathbf{H}|^2}{2} \qquad (15.27)$$

The sum of the diagonal elements, or trace, of this tensor is zero.

Angular Momentum of the Electromagnetic Field. Let us write Eqs. (15.19) for two different spatial indices, α and β, multiply them by x_β and x_α, respectively, and subtract one from the other:

$$x_\beta \frac{\partial T_{\alpha k}}{\partial x_k} - x_\alpha \frac{\partial T_{\beta k}}{\partial x_k} = -4\pi (x_\beta F_{\alpha k} j_k - x_\alpha F_{\beta k} j_k)$$

We transform the left-hand side by parts to get

$$\frac{\partial}{\partial x_k}(x_\beta T_{\alpha k} - x_\alpha T_{\beta k}) - \delta_{k\beta} T_{\alpha k} + \delta_{k\alpha} T_{\beta k}$$

$$= \frac{\partial}{\partial x_k}(x_\beta T_{\alpha k} - x_\alpha T_{\beta k})$$

Thanks to the symmetry of tensor $T_{\alpha\beta}$, the terms outside the derivatives cancel out.

Now integrate both sides of the equation over the three-dimensional volume as was done in obtaining the relationships (15.20a) and (15.20b). Transforming the right-hand side in the same way as before, we reduce it to the form

$$\frac{e}{c}(x_\beta F_{\alpha k}\dot{x}_k - x_\alpha F_{\beta k}\dot{x}_k) = x_\beta \dot{p}_\alpha - x_\alpha \dot{p}_\beta$$

Now, taking into account that $p_\alpha = \dot{x}_\alpha(1 - v^2/c^2)^{-1/2}$ and $p_\beta = \dot{x}_\beta(1 - v^2/c^2)^{-1/2}$, we represent the difference $x_\beta \dot{p}_\alpha - x_\alpha \dot{p}_\beta$ in the form of the total derivative of $x_\beta p_\alpha - x_\alpha p_\beta$, that is, of the angular momentum component of a charged particle with an index not equal to α or β:

$$M_\gamma = \varepsilon_{\gamma\alpha\beta} x_\alpha p_\beta$$

But we then see that on the left we have the time derivative of the angular momentum of the electromagnetic field, which can easily be expressed with the help of (15.27) as

$$\mathbf{M}_{\text{field}} = \int \frac{1}{4\pi c}[\mathbf{r} \times (\mathbf{E} \times \mathbf{H})] \, d^{(3)}\tau \qquad (15.28)$$

It follows from this that the integrand

$$\frac{1}{4\pi c} [\mathbf{r} \times (\mathbf{E} \times \mathbf{H})] \tag{15.29}$$

is the density of the angular momentum of the electromagnetic field.

EXERCISES

1. Defining tensor $F_{ik}^* = \varepsilon_{iklm} F_{lm}$, obtain the equation for it from (15.15). Write the components of F_{ik}^* in matrix form.

2. Develop the energy conservation equation in three-dimensional form from (12.30) and (12.32) (the Poynting theorem).

16

ELECTROSTATICS OF POINT CHARGES

Slowly Variable Fields. An important class of approximate solutions of electrodynamical equations comprises slowly variable fields, for which the time derivatives in Maxwell's equations can be neglected. The remaining terms form two sets of equations, which are entirely independent of each other:

$$\text{div } \mathbf{E} = 4\pi\rho \tag{16.1}$$

$$\text{curl } \mathbf{E} = 0 \tag{16.2}$$

and

$$\text{div } \mathbf{H} = 0 \tag{16.3}$$

$$\text{curl } \mathbf{H} = \frac{4\pi}{c} \mathbf{j} \tag{16.4}$$

The first two equations contain only the electric field and the density of the charge producing the field; the second two equations involve only the magnetic field and current density, the right-hand sides of the equations being regarded as known functions of the coordinates and time. Since there are no time derivatives in (16.2) and (16.4), the time dependence of the electric field is the same as the charge densities, and the time dependence of the magnetic field is the same as the current densities. Hence, to the approximation of (16.1)-(16.4), the field is, as it were, established instantaneously,

in correspondence with the charge and current distribution that generated it.

The fact is that any change in the field is transmitted through space at the speed of light c. If we consider the field at a distance R from a charge, the electromagnetic disturbance will reach it in a time R/c. The charge with a velocity v will be displaced during that time through a distance vR/c. The approximation (16.1)-(16.4) can be applied only when the displacement vR/c does not lead to any essential redistribution of the charge. For example, let a system consist of two equal charges of opposite sign which change places in the time R/c. Then, at the distance R, at the instant $t = R/c$, the electric field will have a direction opposite to the one it would have in the instantaneous propagation at the instant $t = 0$.

Hence, if the dimensions of the system of charges are r and their velocities v (in order of magnitude), then Eqs. (16.1)-(16.4) can be used at a distance from the system for which the inequality $r/v \gg \gg R/c$, or $R \ll rc/v$, is satisfied.

Suppose $v \ll c$. Then the region of applicability of our set of equations will be sufficiently large.

Equations (16.1) and (16.2) are called the *equations of electrostatics*, and (16.3) and (16.4), the *equations of magnetostatics*.

Scalar Potential in Electrostatics. In order to satisfy Eq. (16.2), we put

$$\mathbf{E} = -\text{grad } \varphi$$

According to (12.35), φ is the scalar potential. The equation for the scalar potential is obtained from (16.1)

$$\text{div grad } \varphi = \nabla^2 \varphi = -4\pi\rho \tag{16.6}$$

which also follows from (12.44), if we equate to zero the nonstatic term $(1/c^2)(\partial^2\varphi/\partial t^2)$.

Let us find the solution to Eq. (16.6) for a point charge, that is, we put ρ equal to zero everywhere except at the origin of the coordinate system. Then φ can depend only on the distance from the origin, r.

In Section 11 an expression for the Laplacian was obtained in spherical coordinates (11.46). In the special case, when the required function depends only on r, we obtain from (11.46)

$$\frac{1}{r^2}\frac{d}{dr}r^2\frac{d\varphi}{dr} = -4\pi\rho \tag{16.7}$$

Let us integrate this equation between r_1 and r_2, first multiplying it by r^2. Since the region of integration does not contain the origin, where the point charge is situated, the integral of the right-hand

side becomes zero. Hence

$$r_2^2 \left(\frac{d\varphi}{dr} \right)_2 = r_1^2 \left(\frac{d\varphi}{dr} \right)_1 , \qquad r^2 \frac{d\varphi}{dr} = A = \text{constant}$$

Therefore the potential is

$$\varphi = -\frac{A}{r} + B$$

The constant B is equal to zero if we take the potential to be equal to zero at an infinite distance away from the charge. Let us now determine the constant A. For this, we integrate Eq. (16.6) over a sphere surrounding the origin. Since the Laplacian $\nabla^2\varphi = $ = div grad φ, the space integral can be transformed into an integral over the surface of the sphere. This integral is

$$\int \text{grad } \varphi \, dS = \int \frac{A}{r^2} \, r^2 \, d\Omega = 4\pi A$$

In the right-hand side we have

$$- \int 4\pi\rho \, dV = -4\pi e$$

since the integration region includes the point where the charge is situated, that is, the origin of the coordinate system. Thus $A = -e$. The potential of a point charge is thus

$$\varphi = \frac{e}{r} \tag{16.8}$$

We obtain the same result for a spherically symmetrical volume-charge distribution, if the potential is calculated outside the volume occupied by the charges. In other words, the potential of a charged sphere at all external points is the same as the potential of an equal point charge situated at the centre of the sphere. A similar result is obtained for the gravitational potential, because Newton's gravitational law in form resembles Coulomb law. This fact is used in most astronomical problems, where celestial bodies are considered as gravitating points.

If the origin does not coincide with the charge, and the charge's coordinates are x, y, z, that is, the charge is located at a point with radius vector $\mathbf{r}$, the potential at point X, Y, Z (radius vector $\mathbf{R}$) is

$$\varphi = \frac{e}{|\mathbf{R} - \mathbf{r}|} = \frac{e}{[(X-x)^2 + (Y-y)^2 + (Z-z)^2]^{1/2}}$$

$$= \frac{e}{[(X_\alpha - x_\alpha)(X_\alpha - x_\alpha)]^{1/2}} \tag{16.9}$$

The Potential of a System of Charges. Since the equations of electrodynamics are linear, the potential produced by several charges equals the sum of the potentials of each charge separately. If the

radius vector of the ith charge is $\mathbf{r}^i$, then the total potential of the system is

$$\varphi = \sum_i \frac{e_i}{|\,\mathbf{R}-\mathbf{r}^i\,|} = \sum_i \frac{e_l}{[(X_\alpha - x_\alpha^i)\,(X_\alpha - x_\alpha^i)]^{1/2}}$$

But, to save space, in future we shall write $(X_\alpha - x_\alpha^i)^2$ instead of $(X_\alpha - x_\alpha^i)(X_\alpha - x_\alpha^i)$. Then the potential at a point with radius vector $\mathbf{R}$ is

$$\varphi = \sum_i e_i \, [(X_\alpha - x_\alpha^i)^2]^{-1/2} \qquad (16.10)$$

remembering that inside the brackets we sum over α from 1 to 3.

Suppose now that the origin of a coordinate system is located somewhere inside a domain occupied by charges, for example, the centre of the smallest sphere embracing all the charges. We shall look for the potential at a large distance from the origin, that is, at a distance R for which all the inequalities

$$R \gg r^i \qquad (16.11)$$

are satisfied.

In other words, we must determine the potential at a large distance from a system of charges. For that we should expand function (16.10) in a Taylor series in powers of x_α^i. We shall perform the expansion up to the quadratic term, but shall write it first for only one term of the summation over all the charges, omitting the index i for brevity:

$$[(X_\alpha - x_\alpha)^2]^{-1/2} = [X_\alpha^2]^{-1/2} - x_\beta \frac{\partial}{\partial X_\beta} [X_\alpha^2]^{-1/2}$$
$$+ \frac{1}{2} x_\beta x_\gamma \frac{\partial^2}{\partial X_\beta \, \partial X_\gamma} [X_\alpha^2]^{-1/2} \qquad (16.12)$$

The summation convention permits writing in concise form the Taylor series for a function of several variables. Since $X_\alpha^2 = R^2$, we obtain the expression for the first derivative:

$$\frac{\partial}{\partial X_\beta} [X_\alpha^2]^{-1/2} = \frac{\partial}{\partial X_\beta} \frac{1}{R} = \frac{\partial R}{\partial X_\beta} \frac{\partial}{\partial R} \frac{1}{R} = -\frac{X_\beta}{R^3} \qquad (16.13)$$

where we have used Eq. (11.34), which in the notation of this section, is of the form

$$\frac{\partial R}{\partial X_\beta} = \frac{X_\beta}{R}$$

Thus, the term in the sum (16.12), linear in x_β, is equal to

$$\frac{x_\beta X_\beta}{R^3} = \frac{\mathbf{r}\mathbf{R}}{R^3} \qquad (16.14)$$

It is somewhat more difficult to calculate the term which is quadratic in x_β. We first write the second derivative:

$$\frac{\partial^2}{\partial X_\beta \, \partial X_\gamma} \frac{1}{R} = -\frac{\partial}{\partial X_\gamma} \frac{X_\beta}{R^3} = -\frac{1}{R^3} \frac{\partial X_\beta}{\partial X_\gamma} - X_\beta \frac{\partial}{\partial X_\gamma} \frac{1}{R^3}$$

$$= -\frac{\delta_{\beta\gamma}}{R^3} + X_\beta \frac{\partial}{\partial X_\gamma} \frac{1}{R^3}$$

in accordance with the general definition of tensor $\delta_{\beta\gamma}$ in Section 9. Further

$$\frac{\partial}{\partial X_\gamma} \frac{1}{R^3} = \frac{\partial R}{\partial X_\gamma} \frac{\partial}{\partial R} \frac{1}{R^3} = -\frac{X_\gamma}{R} \frac{3}{R^4} = -\frac{3X_\gamma}{R^5}$$

by the general rule for differentiating a composite function, in this case the function $1/R^3$.

Thus we obtain

$$\frac{\partial^2}{\partial X_\beta \, \partial X_\gamma} \frac{1}{R} = -\frac{\delta_{\beta\gamma}}{R^3} + \frac{3X_\beta X_\gamma}{R^5}$$

Finally, the expansion $|\mathbf{R} - \mathbf{r}|^{-1}$ in powers of components of $\mathbf{r}$ has the form

$$\frac{1}{|\mathbf{R}-\mathbf{r}|} = \frac{1}{R} + \frac{\mathbf{rR}}{R^3} + \frac{1}{2} x_\beta x_\gamma \left(\frac{3X_\beta X_\gamma}{R^5} - \frac{\delta_{\beta\gamma}}{R^3} \right) \qquad (16.15)$$

We now subtract from the last term in (16.15) a quantity identically equal to zero:

$$\frac{1}{6} x_\alpha x_\alpha \delta_{\beta\gamma} \left(\frac{3X_\beta X_\gamma}{R^5} - \frac{\delta_{\beta\gamma}}{R^3} \right) = \frac{1}{6} x_\alpha x_\alpha \left(\frac{3X_\beta X_\beta}{R^5} - \frac{\delta_{\beta\beta}}{R^3} \right) \equiv 0$$

since $X_\beta X_\beta = R^2$, and $\delta_{\beta\beta} = 3$. Then the expansion (16.15) takes the form

$$\frac{1}{|\mathbf{R}-\mathbf{r}|} = \frac{1}{R} + \frac{\mathbf{rR}}{R^3}$$

$$+ \frac{1}{2} \left(x_\beta x_\gamma - \frac{1}{3} \delta_{\beta\gamma} x_\alpha x_\alpha \right) \left(\frac{3X_\beta X_\gamma}{R^5} - \frac{\delta_{\beta\gamma}}{R^3} \right) \qquad (16.16)$$

This series must be substituted into the potential (16.10) and summed over all the charges. We introduce the following abbreviated notation:

$$\mathbf{d} \equiv \sum_i e_i \mathbf{r}^i \qquad (16.17)$$

$$q_{\alpha\beta} \equiv \frac{1}{2} \sum_i e_i \left(x^i_\alpha x^i_\beta - \frac{1}{3} \delta_{\alpha\beta} x^i_\gamma x^i_\gamma \right) \qquad (16.18a)$$

or in terms of the components:

$$q_{xx} = \frac{1}{2} \sum_i e_i \left(x^{i2} - \frac{r^{i2}}{3} \right), \qquad q_{xy} = \frac{1}{2} \sum_i e_i x^i y^i$$

$$q_{yy} = \frac{1}{2} \sum_i e_i \left(y^{i2} - \frac{r^{i2}}{3} \right), \qquad q_{xz} = \frac{1}{2} \sum_i e_i x^i z^i \qquad (16.18b)$$

$$q_{zz} = \frac{1}{2} \sum_i e_i \left(z^{i2} - \frac{r^{i2}}{3} \right), \qquad q_{yz} = \frac{1}{2} \sum_i e_i y^i z^i$$

The vector **d** (the three quantities d_x, d_y, d_z) and the six quantities q_{xx}, q_{yy}, q_{zz}, q_{xy}, q_{xz}, q_{yz} depend only on the charge distribution in the system, and not on the place at which the potential is determined. In the notation of (16.17) and (16.18a) the potential at large distances away from the system is

$$\varphi = \frac{1}{R} \sum e_i + \frac{\mathbf{dR}}{R^3} + q_{\alpha\beta} \left(\frac{3X_\alpha X_\beta}{R^5} \right) \qquad (16.19)$$

since

$$q_{\alpha\beta} \delta_{\alpha\beta} = q_{\alpha\alpha} = \frac{1}{2} \sum_i e_i \left(x^i_\alpha x^i_\alpha - 3 \times \frac{1}{3} x^i_\alpha x^i_\alpha \right) = 0$$

where the terms with different indices, of the type q_{xy}, actually appear twice in the summation (for example, q_{xy} and the equal term q_{yx}).

The vector **d** is called the *dipole moment of a system of charges*; the tensor of rank 2, $q_{\alpha\beta}$, is the *quadrupole moment of the system.*

Dipole Moment. We shall now examine the expression (16.19) obtained for potential. The zero term $\sum e_i/R$ corresponds to the approximation according to which the whole charge is considered to be concentrated at the origin. In other words, it corresponds to a substitution of the entire system of charges by a single point charge.

This approximation is clearly insufficient when the system is neutral, that is, if $\sum e_i = 0$. This case is very usual, since atoms and molecules are neutral (their electronic charge balances the charge of the nuclei).

Let us assume that the total charge is equal to zero and then consider the first term of the expansion involving the dipole moment. This term decreases like R^{-2}, that is, more rapidly than the potential of a charged system. Besides, it is proportional to the cosine of the angle between **d** and **R**. The simplest way to produce a neutral system is by taking two equal and opposite charges. Such a system

is called a *dipole*. Its moment is

$$\mathbf{d} = \sum_i e_i \mathbf{r}^i = e\,(\mathbf{r}^+ - \mathbf{r}^-) \tag{16.20}$$

in accordance with the definition used in general courses of physics that the dipole moment is the product of the charge by the vector drawn from the positive to the negative charge.

It can be seen from Eq. (16.20) that the definition of dipole moment does not depend on the choice of coordinate origin, since it involves only the relative position of the charges. We shall show that the dipole moment always possesses this property, provided the total charge of the system is zero.

Indeed, if we displace the origin through some distance **a**, then the radius vectors of all the charges change to

$$\mathbf{r}^i = \mathbf{r}'^i + \mathbf{a}$$

Substituting this into the expression for the dipole moment, we obtain

$$\mathbf{d} = \sum_i e_i \mathbf{r}^i = \sum_i e_i \mathbf{r}'^i + \mathbf{a} \sum_i e_i = \sum_i e_i \mathbf{r}'^i = \mathbf{d}' \tag{16.21}$$

because $\sum_i e_i = 0$.

But if the system is not neutral, then we choose **a** in the following manner:

$$\mathbf{a} = \left(\sum_i e_i \mathbf{r}^i\right)\left(\sum_i e_i\right)^{-1} \tag{16.22}$$

This choice is analogous to the choice of centre of mass for a system of masses. Thus, we can say that in a system which is not as a whole neutral the vector **a** determines the *electrical centre* of the system of charges. For a neutral system it is impossible to determine **a**, since the denominator of (16.22) is zero. If for a charged system we choose **a** according to (16.22), then $\sum_i e_i \mathbf{r}'^i = 0$, that is, the dipole moment of a charged system relative to its electric centre is equal to zero.

We thus have the following alternatives: either the system is neutral, and then the expansion (16.19) begins with a dipole term independent of the choice of coordinate origin, or $e = \sum_i e_i$ is a resultant charge, and then the dipole term in the expansion is equal to zero for a corresponding choice of origin.

Quadrupole Moment. In the expansion (16.19) we now consider the second term containing the quadrupole moment. A quadrupole is a system of two dipoles of moment **d**, which are equal in magnitude and opposite in direction. It is clear that a potential expansion for such a system will have neither a zero nor a first term, so that Eq. (16.19) contains only a second term on the right-hand side.

The simplest quadrupole can be formed by placing four charges at the vertices of a parallelogram, where the charges are of equal magnitude but with pairs of charges having opposite signs. The charges alternate when we traverse the vertices of the parallelogram. Such a system is neutral. However, a charged system, too, can have a quadrupole moment. It indicates to what extent the charge distribution in the system differs from spherical symmetry.

Indeed, in this section it was shown that the potential due to a spherically symmetrical system of charges decreases in strict accordance with a R^{-1} law, and the potential due to a quadrupole follows a R^{-3} law. For this reason, the quadrupole term in the potential expansion can arise only in the case of a nonspherical charge distribution. The subsequent expansion terms, which can be obtained in the same way, take account of increasingly fine deviations from spherical symmetry in the charge distribution.

Let us now determine in what sense the quadrupole moment characterizes a nonspherical distribution. Equation (16.22) establishes the analogy between the centre of mass of a system of masses and the electric centre of a system of charges. Similarly, Eq. (16.18a) allows us to establish a certain correspondence between the components of a quadrupole moment and the moments of inertia of the system of masses defined by Eqs. (9.3) and (9.4).

Since we are concerned with the similarity between, not identity of, quantities, we can disregard the fact that (16.18a) involves a summation, while (9.3) involves an integration. Besides, this difference disappears if we take a continuous charge distribution or discrete mass distribution (as of nuclei in a molecule). Furthermore, we shall forget for the moment that moment-of-inertia components involve masses, not charges.

The tensor expression for moment of inertia has the form

$$I_{\alpha\beta} = \sum_m m \left(\delta_{\alpha\beta} x_\gamma x_\gamma - x_\alpha x_\beta \right)$$

We put $\alpha = \beta$ and perform the summation according to the general rule. We then obtain

$$I_{\alpha\alpha} = \sum_m m \left(3 x_\alpha x_\alpha - x_\alpha x_\alpha \right) = 2 \sum_m m x_\alpha x_\alpha$$

We substitute $\sum_m m x_\alpha x_\alpha$ into the initial expression to get

$$\sum_m m x_\alpha x_\beta = \frac{1}{2} \delta_{\alpha\beta} I_{\gamma\gamma} - I_{\alpha\beta}$$

Knowing how $\sum m x_\alpha x_\alpha$ and $\sum m x_\alpha x_\beta$ are expressed in terms of $I_{\alpha\beta}$, we substitute them into the definition of a quadrupole moment:

$$q_{\alpha\beta} \sim \frac{1}{2} \left(\frac{1}{3} \delta_{\alpha\beta} I_{\gamma\gamma} - I_{\alpha\beta} \right) \qquad (16.23a)$$

The symbol $\sim$ serves as a reminder that what we have is only a correspondence.

In Section 9 it was shown that moments of inertia can be reduced to the principal axes, that is, a coordinate system can be found in which the products of inertia vanish, leaving only the diagonal elements of the inertia tensor. But, since the relationships between $q_{\alpha\beta}$ and $I_{\alpha\beta}$ exist in any coordinate system, for the same principal axes, the quadrupole moment components with different indices also vanish. The quadrupole moment relative to the principal axes is made to correspond with the moment of inertia as follows:

$$q_1 \sim \frac{1}{6}\,(I_2 + I_3 - 2I_1) \tag{16.23b}$$

(and for other components similarly).

If the system possesses spherical symmetry, then $I_1 = I_2 = I_3$, so that $q_1 = q_2 = q_3 = 0$. Therefore, the presence of a quadrupole moment in a system of charges indicates that the charge distribution is not spherically symmetrical. However, a reverse assertion would not be true: if the quadrupole moment is equal to zero, the system of charges may not be spherically symmetrical. It is necessary, in expansion (16.19), to take into account terms of higher order than written here, and only if they are all equal to zero do we have spherical symmetry. Only then does the potential decrease strictly as R^{-1}.

It will be noted that from the definition of a quadrupole moment, or from (16.23b), there follows directly the identity $q_{\alpha\alpha} = 0$, $q_1 + q_2 + q_3 = 0$, that is, only two of the three principal components of a quadrupole moment are independent.

The relations (16.23a) and (16.23b) should be regarded literally if we are talking about a gravitational potential. We know that the earth is not strictly spherical, but is flattened at the poles. Therefore, the force of gravity contains terms that decrease faster than the inverse square law. This affects the motion of the moon and, even more so, that of artificial satellites, which are closer to the earth. For them Eq. (16.19) would have to be written up to a higher approximation.

Equations (16.23b) become simpler if two moments of inertia of the system are equal, that is, if the system's symmetry requires the equality $I_1 = I_2$. Then

$$q_1 \sim \frac{1}{6}\,(I_3 - I_1) \equiv -\frac{q}{2}$$

$$q_2 \sim \frac{1}{6}\,(I_3 - I_1) \equiv -\frac{q}{2}$$

$$q_3 \sim \frac{1}{3}\,(I_1 - I_3) \equiv q$$

In this case the quadrupole moment has only one independent component q. Its sign is called the sign of the quadrupole moment. The quantity $q = \sum e_i \, (z^{i2} - r^{i2}/3)$.

If the charges were distributed with spherical symmetry, we would have the equality $\sum e_i r^{i2} = 3 \sum e_i z^{i2}$, for then $\sum e_i x^{i2} = \sum e_i y^{i2} = \sum e_i z^{i2}$ for any choice of axes. Then, obviously, q, too, would be equal to zero.

The positive sign of q shows that $\sum e_i z^{i2} > \sum e_i r^{i2}/3$, that is, it indicates a charge distribution extending along the z axis; a negative sign indicates a flattened charge distribution.

From (16.19), the potential due to such a quadrupole with one component q is

$$
\begin{aligned}
\varphi_q &= -\frac{1}{2} q \left(\frac{3X^2}{R^5} - \frac{1}{R^3} \right) - \frac{1}{2} q \left(\frac{3Y^2}{R^5} - \frac{1}{R^3} \right) + q \left(\frac{3Z^2}{R^5} - \frac{1}{R^3} \right) \\
&= -\frac{3}{2} q \left(\frac{X^2 + Y^2 - 2Z^2}{R^5} \right) = -\frac{3}{2} q \left(\frac{R^2 - 3Z^2}{R^5} \right) \\
&= -\frac{3}{2} \frac{q}{R^3} (1 - 3 \cos^2 \vartheta)
\end{aligned} \tag{16.24}
$$

The potential of such a quadrupole depends on the angle ϑ according to the law $1 - 3 \cos^2 \vartheta$, where ϑ is the angle between the axis of symmetry of the quadrupole and the radius vector of the point at which the potential is determined. Such deviations from spherical symmetry have been found in the electrostatic potential of many nuclei. The quadrupole moments of nuclei give us an insight into their structure.

The Energy of a System of Charges in an Electrostatic Field. We shall now calculate the energy of a system of charges in an external electric field. The potential energy of a charge in a field is equal to $U = e\varphi$, because the force acting on the charge is equal to $\mathbf{F} = -\operatorname{grad} U = -e \operatorname{grad} \varphi = e\mathbf{E}$. The energy of a system of charges is thus

$$
U = \sum_i e_i \varphi \, (\mathbf{r}^i) \tag{16.25}
$$

where $\mathbf{r}^i$ is the radius vector for the ith charge.

Let us suppose that the field does not change much over the space occupied by the charges, so that the potential at the site of the ith charge can be expanded in a Taylor series:

$$
\varphi \, (\mathbf{r}) = \varphi \, (0) + x_\alpha \cdot \left(\frac{\partial \varphi}{\partial x_\alpha} \right)_0 + \frac{1}{2} x_\alpha x_\beta \left(\frac{\partial^2 \varphi}{\partial x_\alpha \, \partial x_\beta} \right)_0 + \dots \tag{16.26}
$$

We transform the last term in the same way as in the expansion (16.15), taking advantage of the fact that φ is the potential of the

external field (and not the field produced by the given charges), so that $\nabla^2\varphi = 0$. We subtract from φ the quantity $(r^2/6)\nabla^2\varphi$ equal to zero. Then, after summation over the charges, we obtain

$$U = \varphi(0) \sum_i e_i - (\mathbf{d} \cdot \mathbf{E_0})$$

$$+ \frac{1}{2} \sum_i e_i \left(x_\alpha^i x_\beta^i - \frac{1}{3} \delta_{\alpha\beta} x_\gamma^i x_\gamma^i \right) \left(\frac{\partial^2\varphi}{\partial x_\alpha \, \partial x_\beta} \right)_0 + \cdots$$

$$= \varphi(0) \sum_i e_i - (\mathbf{d} \cdot \mathbf{E_0}) + q_{\alpha\beta} \left(\frac{\partial^2\varphi}{\partial x_\alpha \, \partial x_\beta} \right)_0 + \cdots \qquad (16.27)$$

Here, the value of the field $(\text{grad } \varphi)_0 = -\mathbf{E_0}$ at the origin has been substituted into the term involving the dipole moment. Relating Eq. (16.27) to the principal axes of the quadrupole moment, we can rewrite it as follows:

$$U = \varphi(0) \sum_i e_i - (\mathbf{d} \cdot \mathbf{E_0})$$

$$+ \left((q_1 + q_2) \frac{\partial E_3}{\partial x_3} - q_1 \frac{\partial E_1}{\partial x_1} - q_2 \frac{\partial E_2}{\partial x_2} \right) \qquad (16.28)$$

In the case of a neutral system, the term involving dipole moment is especially important. The quadrupole term accounts for the extension of the system, since it involves field derivatives. If the system is spherically symmetrical, that is, if it has a quadrupole moment equal to zero, there is no correction for extension. Higher order corrections are also absent, so that the potential energy will always depend only on the value of the potential at the centre. This is why spherical bodies not only attract, but are also attracted, as points Of course these assertions are mutually related by Newton's Third Law, which holds for electrostatics, since fields are determined by the instantaneous configuration of charges.

EXERCISES

1. Show that the mean value of the potential over a spherical surface is equal to its value at the centre of the sphere, if the equation $\nabla^2\varphi = 0$ is satisfied over the whole volume. Relate this to the result obtained for the potential energy of a spherically symmetrical system of charges in an external field.

Hint. The potential should be expanded in a series involving the powers of the sphere's radius. In integration over the surface, all the terms containing x, y, and z an odd number of times become zero. The terms containing

x, y, and z an even number of times can be rearranged so that they are proportional to $\nabla^2\varphi$, $\nabla^2\nabla^2\varphi$, and so on. There remains only the zero term of the expansion, which proves the theorem.

2. Calculate the electric field of a dipole and the interaction energy of two dipoles located at a large distance from each other.

3. Reduce to quadratures the problem on the motion of a charged particle in a dipole field in the nonrelativistic approximation.

Hint. Use the Hamilton-Jacobi equation and separate the variables. The integral over the angle cannot be expressed in elementary form.

17

MAGNETOSTATICS OF POINT CHARGES

The Meaning of the Equations of Magnetostatics. In the preceding section it was shown that if the velocities of charges are small in comparison with the speed of light, then the magnetic field satisfies the set of equations (see (16.3) and (16.4)):

$$\text{div } \mathbf{H} = 0 \tag{17.1}$$

$$\text{curl } \mathbf{H} = \frac{4\pi}{c}\, \mathbf{j} = \frac{4\pi}{c}\, \rho\mathbf{v} \tag{17.2}$$

They are called the *equations of magnetostatics*. However, it is not hard to see that they cannot be directly satisfied for point charges. For this we take the divergence of both sides of (17.2). On the left we have identically div curl $\mathbf{H} = 0$, while on the right we obtain $(4\pi/c)$ div $\rho\mathbf{v}$. For point charges it is impossible to make div $\rho\mathbf{v}$ equal to zero simultaneously at all points of space, because there are no closed current lines. Closed lines can appear only when the paths of the charges, that is, their motions over a specified time, are considered. Therefore Eq. (17.2) has no meaning for instantaneous values and becomes meaningful only for values averaged over time.

If the time averaged value is determined from (10.48), it is not necessary to consider the charge paths to be closed in the strict sense; it is sufficient for them to be finite, or even for the averaged quantity integrated over time to increase not faster than the time itself.

Along the lines of (10.48), we define the mean of a certain function averaged over time by

$$\overline{f} = \frac{1}{t_0} \int\limits_{0}^{t_0} f(\mathbf{r},\,\mathbf{v})\, dt \tag{17.3}$$

This averaging operation is commutative with a differentiation of the function with respect to the coordinates, since it is performed with respect to a different variable, time.

Let us perform the averaging on the quantity div **j**. Taking advantage of the commutativity of the operation div curl, that is, of the determination of the partial derivatives with respect to the coordinates, with integration over time, we rewrite Eq. (17.2) in the mean form:

$$\text{div curl}\ \overline{\mathbf{H}} = \frac{1}{t_0} \int_0^{t_0} \text{div}\ \mathbf{j}\, dt = -\frac{1}{t_0} \int_0^{t_0} \frac{\partial \rho}{\partial t}\, dt :$$

$$= -\frac{\rho(t_0) - \rho(0)}{t_0} \tag{17.4}$$

Suppose now that the difference $\rho(t_0) - \rho(0)$ increases slower than the time interval t_0 itself. This is valid a priori for any quasiperiodic motion (see Sec. 10). If t_0 is chosen sufficiently large, the ratio $[\rho(t_0) - \rho(0)]/t_0$ becomes infinitesimal in the limit. Thanks to that, the mean value of the current indeed satisfies the equation

$$\frac{1}{t_0} \int_0^{t_0} \text{div}\ \mathbf{j}\, dt = \frac{1}{t_0}\ \text{div} \int_0^{t_0} \mathbf{j}\, dt = \text{div}\ \overline{\mathbf{j}} = 0 \tag{17.5}$$

Hence Eq. (17.2) and all subsequent equations of this section should be understood in the sense of time averages, which will be denoted by bars over the quantities referring to the motion of charges. We shall further agree not to draw a bar over **H**, though it is implied.

If the condition $\overline{(\partial f/\partial t)} = 0$ is satisfied not only for the charge density, but for all functions relating to the motion of the charges, such motion is said to be *stationary* or *steady*.

A special case of steady motion is periodic motion, for example, uniform circular motion. But for a steady state it is sufficient for the charges simply to be moving in a limited region or to be receding at a rate slower than time increases.

The Vector Potential Equations. In order to satisfy Eq. (17.1), we put, as in the most general case,

$$\mathbf{H} = \text{curl}\ \mathbf{A} \tag{17.6}$$

(cf. (12.34)); here **A** is the vector potential. Equation (17.6) does not fully define **A**, because if we add to **A** the gradient of an arbitrary function f, as in (12.36), the expression for curl **A** will not change. Therefore an additional condition must be imposed on **A**. The Lorentz condition (12.42) suggests that we should require

$$\text{div}\ \mathbf{A} = 0 \tag{17.7}$$

Then, substituting into (17.6) and (17 2), we obtain

$$\text{curl } \mathbf{H} = \text{curl curl } \mathbf{A} = \frac{4\pi}{c} \bar{\mathbf{j}} \qquad (17.8)$$

But from (11.42)

$$\text{curl curl } \mathbf{A} = \text{grad div } \mathbf{A} - \nabla^2\mathbf{A} = -\nabla^2\mathbf{A} \qquad (17.9)$$

where we have made use of condition (17.7). Hence, $\mathbf{A}$ satisfies the equation

$$\nabla^2\mathbf{A} = -\frac{4\pi}{c}\bar{\mathbf{j}} = -\frac{4\pi}{c}\overline{\rho\mathbf{v}} \qquad (17.10)$$

which is entirely analogous to (16.6) for a point charge.

Equation (17.10) can also be obtained from (12.43), if we discard the term $(1/c^2)(\partial^2\mathbf{A}/\partial t^2)$, which is inessential in magnetostatics.

Since the solution of (17.10) looks exactly like the solution of (16.6), let us write it in a form analogous to (16.9) for a separate point charge. Each component of $\mathbf{A}$ satisfies an equation of the form (16.6), with the difference that in the right-hand side we have the vector $\mathbf{j}$ (j_x, j_y, j_z). If (17.10) is expanded in components in Cartesian coordinates, we obtain three equations of the form (16.6) having in the right-hand sides $-(4\pi/c)\overline{\rho v_x}$, $-(4\pi/c)\overline{\rho v_y}$, and $-(4\pi/c)\overline{\rho v_z}$, respectively. Hence the vector potential of a point charge is

$$\mathbf{A} = \frac{e\mathbf{v}}{c\,|\mathbf{R}-\mathbf{r}|} \qquad (17.11)$$

We shall now show that $\mathbf{A}$ satisfies condition (17.7). The divergence must be taken with respect to the radius vector at the point at which $\mathbf{A}$ is determined.

But $\text{grad}_{\mathbf{R}}\,|\mathbf{R}-\mathbf{r}|^{-1} = -\text{grad}_{\mathbf{r}}\,|\mathbf{R}-\mathbf{r}|^{-1}$, so that

$$\text{div } \mathbf{A} = \frac{e}{c}\,\overline{\mathbf{v}\,\text{grad}_{\mathbf{R}}\,\frac{1}{|\mathbf{R}-\mathbf{r}|}} = -\frac{e}{c}\,\overline{\mathbf{v}\,\text{grad}_{\mathbf{r}}\,\frac{1}{|\mathbf{R}-\mathbf{r}|}}$$

$$= -\frac{e}{c}\,\overline{\frac{d}{dt}\,\frac{1}{|\mathbf{R}-\mathbf{r}|}}$$

The expression in the right-hand side is the total time derivative of the quantity $|\mathbf{R}-\mathbf{r}|^{-1}$. From the steady-state condition, it is equal to zero.

We shall now calculate the mean magnetic field of a point charge. Using Eq. (11.28), we obtain

$$\mathbf{H} = \text{curl } \mathbf{A} = \frac{e}{c}\left(\text{grad}_{\mathbf{R}}\,\frac{1}{|\mathbf{R}-\mathbf{r}|}\times\mathbf{v}\right)$$

$$= -\frac{e}{c\,|\mathbf{R}-\mathbf{r}|^3}\,\mathbf{v}\times(\mathbf{R}-\mathbf{r}) \qquad (17.12)$$

This equation refers, of course, only to steady motion. In particular it is applied to direct current.

Vector Potential at Great Distances From a System of Stationary Currents. The vector potential for a system of point charges is equal to the sum of the vector potentials for each charge separately:

$$\mathbf{A} = \sum_i \frac{e_i \mathbf{v}^i}{c \,|\mathbf{R} - \mathbf{r}^i|} \tag{17.13}$$

We shall now obtain approximate formulas valid at great distances from the system, similar to those obtained in electrostatics. For this we substitute into (17.13) the expansion (16.15), in which only the first term linear with respect to $\mathbf{r}^i$ has been retained:

$$\frac{1}{|\mathbf{R} - \mathbf{r}^i|} = \frac{1}{R} + \frac{\mathbf{r}^i \mathbf{R}}{R^3} \tag{17.14}$$

The vector potential in the approximation (17.14) is of the form

$$\mathbf{A} = \frac{1}{Rc} \overline{\sum_i e_i \mathbf{v}^i} + \frac{1}{cR^3} \sum_i \overline{e_i \,(\mathbf{R} \cdot \mathbf{r}^i)\, \mathbf{v}^i}$$

$$= \frac{1}{Rc} \overline{\frac{d}{dt} \sum_i e_i \mathbf{r}^i} + \frac{1}{cR^3} \sum_i \overline{e_i \,(\mathbf{R} \cdot \mathbf{r}^i)\, \mathbf{v}^i} \tag{17.15}$$

since $\dot{\mathbf{r}}^i \equiv \mathbf{v}^i$. The zero term of the expansion is a total time derivative and vanishes after averaging. We now transform the first term of the expansion, using the identity

$$0 = \overline{\frac{d}{dt} \sum_i e_i \,(\mathbf{R} \cdot \mathbf{r}^i)\, \mathbf{r}^i}$$

$$= \sum_i \overline{e_i \,(\mathbf{R} \cdot \mathbf{r}^i)\, \mathbf{v}^i} + \sum_i \overline{e_i \,(\mathbf{R} \cdot \mathbf{v}^i)\, \mathbf{r}^i} \tag{17.16}$$

From this it follows that into (17.15) we can substitute half the difference of the expressions in the right-hand side of (17.16). Then the vector potential will be

$$\mathbf{A} = \frac{1}{2R^3 c} \sum_i e_i \overline{[\mathbf{v}^i \,(\mathbf{r}^i \cdot \mathbf{R}) - \mathbf{r}^i \,(\mathbf{v}^i \cdot \mathbf{R})]}$$

$$= -\frac{1}{2R^3 c} \sum_i \overline{e_i \mathbf{R} \times (\mathbf{r}^i \times \mathbf{v}^i)} \tag{17.17}$$

We now interchange the signs of the summation and vector product and obtain the required equation:

$$\mathbf{A} = -\frac{\mathbf{R}}{R^3} \times \sum_i \frac{e_i}{2c} \overline{(\mathbf{r}^i \times \mathbf{v}^i)} \tag{17.18}$$

The quantity appearing under the summation sign in (17.18) depends only on the mean current distribution in the system:

$$\bar{\bar{\mu}} \equiv \sum_i \frac{e_i}{2c} \overline{(\mathbf{r}^i \times \mathbf{v}^i)} \qquad (17.19)$$

It is called the *mean magnetic moment of a system of currents*. Using the notation (17.19), we rewrite the vector potential at a large distance from a system of currents in the following form:

$$\mathbf{A} = -\frac{\mathbf{R}}{R^3} \times \bar{\mu} = \operatorname{grad} \frac{1}{R} \times \bar{\mu} \qquad (17.20)$$

Let us now calculate the magnetic field from the vector potential. By definition

$$\mathbf{H} = \operatorname{curl} \mathbf{A} = \operatorname{curl} \left(\operatorname{grad} \frac{1}{R} \times \bar{\mu} \right)$$

Since $\bar{\mu}$ is a constant vector, Eq. (11.30) gives

$$\mathbf{H} = (\bar{\mu} \cdot \nabla) \operatorname{grad} \frac{1}{R} - \bar{\mu} \nabla^2 \frac{1}{R} = -(\bar{\mu} \cdot \nabla) \frac{\mathbf{R}}{R^3}$$

because $\nabla^2 R^{-1} = 0$. Further, $(\bar{\mu} \cdot \nabla) \mathbf{R} = \bar{\mu}$ (see (11.36)). In order to calculate $(\bar{\mu} \cdot \nabla) R^{-3}$, we use Eq. (11.34). This yields

$$(\bar{\mu} \cdot \nabla) \frac{1}{R^3} = \left(\bar{\mu} \cdot \operatorname{grad} \frac{1}{R^3} \right) = -\frac{3}{R^4} (\bar{\mu} \cdot \operatorname{grad} R) = -\frac{3(\bar{\mu} \cdot \mathbf{R})}{R^5}$$

Finally, collecting both terms, we arrive at an equation for $\mathbf{H}$:

$$\mathbf{H} = \frac{3\mathbf{R}(\mathbf{R} \cdot \bar{\mu}) - R^2 \bar{\mu}}{R^5} \qquad (17.21)$$

For comparison, we present the expression for the electric field of a dipole:

$$\mathbf{E} = -\operatorname{grad} \varphi = -\operatorname{grad} \frac{(\mathbf{R} \cdot \mathbf{d})}{R^3} = \frac{3\mathbf{R}(\mathbf{R} \cdot \mathbf{d}) - R^2 \mathbf{d}}{R^5} \qquad (17.22)$$

Thus, both expressions for physically observable quantities, that is, electric and magnetic fields, are entirely analogous. The electric and magnetic moments determine the corresponding fields in a similar way. This explains the name "magnetic moment".

In the case of a charge moving in a plane closed orbit, the definition of magnetic moment (17.19) coincides with the elementary definition of moment in terms of "magnetic sheet". The vector product $\mathbf{r} \times \mathbf{v}$ is twice the area swept out by the radius vector of the charge in unit time (see Section 5, following Eq. (5.4)). Hence $\mathbf{r} \times \mathbf{v} = 2d\mathbf{S}/dt$. By definition of the mean value (17.3)

$$\bar{\mu} = \frac{1}{t_0} \int_0^{t_0} \frac{e}{c} \frac{d\mathbf{S}}{dt} \, dt = \frac{e}{ct_0} \mathbf{S} \qquad (17.23)$$

Here, t_0 is the time of orbital revolution of the charge. In this time the charge passes every point on the orbit once; hence the mean current is equal to $I = e/t_0$. This yields the elementary definition of a magnetic moment:

$$\bar{\mu} = \frac{IS}{c} \tag{17.24}$$

The similarity of Eqs. (17.21) and (17.22) is proof of the equivalence of a closed current, that is, a "magnetic sheet", and a fictitious dipole with the same moment $\bar{\mu}$. At a large distance from a system of currents the field is produced as it were by a dipole.

A System of Moving Point Charges in an External Magnetic Field.
The approximation of slowly varying fields examined in this section holds only when the charge velocities are small in comparison with the speed of light. It is of interest to investigate the motion of such charges in an external magnetic field. For this we must develop a nonrelativistic approximation of the Lagrangian of a system of moving charges.

If in a nonrelativistic approximation we substitute the kinetic energy $m|\mathbf{v}|^2/2$ for $-mc^2(1 - |\mathbf{v}|^2/c^2)^{1/2}$ in the exact Lagrangian (14.23), the latter acquires the following form for a system of identical particles ($e_i = e$, $m_i = m$):

$$L = \sum_i \frac{m|\mathbf{v}_i|^2}{2} + \sum_i \left[\frac{e}{c} (\mathbf{A}(\mathbf{r}_i \cdot \mathbf{v}_i) - e\varphi(\mathbf{r}_i) \right] \tag{17.25}$$

Owing to the term linear with respect to the velocities, in this approximation, too, L is not represented as the difference between the kinetic and potential energies, $L = T - U$. But the Lagrangian can be reduced to its usual form by means of a corresponding change in the reference frame.

We assume the external field to be uniform: $\mathbf{H} = $ constant. The vector potential of such a field is conveniently represented in the form

$$\mathbf{A} = \frac{1}{2}(\mathbf{H} \times \mathbf{r}) \tag{17.26}$$

since, from (11.30),

$$\mathrm{curl}\, \frac{1}{2}(\mathbf{H} \times \mathbf{r}) = \frac{1}{2}[\mathbf{H}\, \mathrm{div}\, \mathbf{r} - (\mathbf{H} \cdot \nabla)\mathbf{r}] = \mathbf{H}$$

Substituting this expression into the Lagrangian and performing a cyclic permutation in the mixed vector product, we obtain

$$L = \sum_i \frac{m|\mathbf{v}_i|^2}{2} + \sum_i \frac{e}{2c}(\mathbf{r}_i \times \mathbf{v}_i)\,\mathbf{H} - e\varphi \tag{17.27}$$

Compare this expression with the Lagrangian (8.5) for the motion of material points with respect to a rotating reference frame.

If we assume the angular velocity of the system to be small, so that its square can be legitimately neglected in comparison with the terms involving it linearly, the Lagrangian (8.5) for a system of identical particles can be rewritten as follows (after the cyclic permutation employed in writing (17.27)):

$$L = \sum_i \frac{m |v_i|^2}{2} + \sum_i m (r_i \times v_i) \omega - U \tag{17.28}$$

Comparing (17.27) and (17.28), we find that the motion of a system of charges in an external constant and uniform magnetic field coincides with its motion relative to a reference frame rotating with a constant angular velocity

$$\omega = \frac{eH}{2mc} \tag{17.29}$$

This assertion is known as *Larmor's theorem.*

We assume the frequency ω to be much smaller than all the frequencies characterizing the proper motions of the charges in the same system in the absence of an external magnetic field. This motion remains unchanged after the magnetic field is turned on, if we go over to a reference frame revolving with the angular velocity $-\omega$. The terms of the Lagrangian linear with respect to the velocity and the terms due to the magnetic field and the rotation mutually cancel out. This, of course, is valid only in the linear approximation, when the effect of the centrifugal force, which is quadratic with respect to the angular velocity, can be neglected. This is equivalent to the assumption of the smallness of ω in comparison with the proper frequencies of the system.

Let a system of like charges possess a magnetic moment defined in accordance with (17.19). Taking the charge e outside the summation sign, we observe that in such a system the magnetic moment is proportional to the angular momentum (see (4.20)):

$$\mu = e \sum \frac{1}{2c} (r_i \times v_i) = \frac{e}{2mc} M \tag{17.30}$$

In such a system the magnetic moment and the angular momentum are conserved together. In an external magnetic field, the system's rotation about the field is of the same type as the rotation of a free symmetric top about the direction of the total angular momentum. For that reason, the motion of a system possessing magnetic moment in an external magnetic field is, like the similar motion of a top, called precession. It takes place with a frequency defined from (17.29), and is accordingly known as *Larmor's precession.*

Let us write the equation for this precession. Whereas in the absence of an external magnetic field the angular momentum vector **M** was constant relative to fixed axes, according to Larmor's theorem, in a field it is constant relative to axes rotating with the angular velocity $-e\mathbf{H}/(2mc)$. The components of **M** along these axes are constant in time ($\dot{\mathbf{M}} = 0$). To go over to fixed axes we should make use of Eq. (9.14), assuming them to be rotating in the opposite sense of the fixed axes with an angular velocity $e\mathbf{H}/(2mc)$. Hence the projections of the moment on the fixed axes are not constant, and the equation for it is

$$\frac{d\mathbf{M}}{dt} + \boldsymbol{\omega} \times \mathbf{M} = 0$$

Substituting $\boldsymbol{\omega}$ according to Larmor's theorem, we obtain

$$\frac{d\mathbf{M}}{dt} = -\frac{e}{2mc}\mathbf{H} \times \mathbf{M} = \boldsymbol{\mu} \times \mathbf{H} \tag{17.31a}$$

or

$$\frac{d\boldsymbol{\mu}}{dt} = \frac{e}{2mc}\boldsymbol{\mu} \times \mathbf{H} \tag{17.31b}$$

If in a system without a magnetic field the angular momentum is not conserved, then (17.31b) can be referred to $\bar{\boldsymbol{\mu}}$.

By first multiplying Eq. (17.31b) scalarly by $\boldsymbol{\mu}$, we find that $d\mu^2/dt = 0$, so that $\mu^2 =$ constant. We then multiply scalarly by **H** and find that $(d/dt)(\boldsymbol{\mu}\cdot\mathbf{H}) = \mu\,|\,\mathbf{H}\,|\,(d/dt)\cos\vartheta = 0$, ($\vartheta$ is the angle between $\boldsymbol{\mu}$ and **H**). Thus, the magnetic moment retains its absolute value and rotates about the magnetic field at a constant angle to it. It is this motion that is called precession.

Note that a system of like charges, the resultant magnetic moment of which is zero, also begins to rotate in a magnetic field.

We shall now pass from the Lagrangian (17.27) to the Hamiltonian of the system. From (14.25) it follows that if the energy of a charged particle is expressed in terms of its velocity, then the magnetic field is eliminated, leaving only the scalar potential. But since expression (14.24), which relates the velocity and momentum, involves a vector potential, the Hamiltonian involves the magnetic field. Substituting the velocity expressed in terms of the momentum instead of the kinetic energy, we obtain the Hamiltonian:

$$\mathscr{H} = \frac{1}{2m}\sum_i \left(\mathbf{p}_{0i} - \frac{e}{c}\mathbf{A}\,(\mathbf{r}_i)\right)^2 + \sum_i e\varphi\,(\mathbf{r}_i) \tag{17.32}$$

Assuming the magnetic field weak, and accordingly neglecting the square of the vector potential, we write the Hamiltonian as

$$\mathscr{H} = \frac{1}{2m}\sum_i |\,\mathbf{p}\,|_{0i}^2 + \frac{e}{cm}\sum_i [(\mathbf{p}_{0i}\cdot\mathbf{A}\,(\mathbf{r}_i)) + e\varphi\,(\mathbf{r}_i)]$$

Since the magnetic field is already involved linearly in the term containing the product $(\mathbf{p}_{0i} \cdot \mathbf{A}(\mathbf{r}_i))$, $\mathbf{p}_{0i}$ must simply be replaced by $m\mathbf{v}_i$. After a cyclic permutation of vectors we obtain

$$\mathscr{H} = \frac{1}{2m} \sum_i |\mathbf{p}|_{0i}^2 - (\mathbf{H} \cdot \boldsymbol{\mu}) + \sum_i e\varphi(\mathbf{r}_i) \qquad (17.33)$$

Thus, in a magnetic field the Hamiltonian acquires a supplementary term $-(\mathbf{H} \cdot \boldsymbol{\mu})$ similar to $-(\mathbf{E} \cdot \mathbf{d})$ in an electric field.

Suppose now that the external magnetic field is not constant and is weakly nonhomogeneous in space, so that its variation over a distance of the order of the dimensions of the system of charges is not great. It then follows from the Hamiltonian (17.33) that[5]

$$\sum_i \dot{\mathbf{p}}_{0i} \equiv \mathbf{F} = -\frac{\partial \mathscr{H}}{\partial \mathbf{r}} = \frac{\partial}{\partial \mathbf{r}} (\mathbf{H} \cdot \boldsymbol{\mu}) \equiv \operatorname{grad}(\mathbf{H} \cdot \boldsymbol{\mu}) \qquad (17.34)$$

A system possessing magnetic moment is subject to a force dependent on the nonuniformity of the field. The equation for this force can be transformed in the following way. Expanding (17.34) with the help of (11.32), we obtain

$$\mathbf{F} = (\boldsymbol{\mu} \cdot \nabla) \mathbf{H} + \boldsymbol{\mu} \times \operatorname{curl} \mathbf{H}$$

But for an external field curl $\mathbf{H}$ is zero, so that the force acting on a system of charges possessing magnetic moment is

$$\mathbf{F} = (\boldsymbol{\mu} \cdot \nabla) \mathbf{H} \qquad (17.35)$$

It manifests itself in the attraction of magnetized bodies to the poles of magnets, where the magnetic field is stronger. A similar expression for force is obtained for a system with an electric dipole moment.

EXERCISES

1. In the expansion of a vector potential, find the term due to the magnetic quadrupole moment.

Solution. Write the factor in the second term of the expansion of $|\mathbf{R} - \mathbf{r}|^{-1}$ due to one charge:

$$(\mathbf{r} \cdot \nabla)^2 \frac{1}{R} = \frac{3(\mathbf{r} \cdot \mathbf{R})^2 - r^2 R^2}{R^5}$$

Then transform the products by parts:

$$\mathbf{v}(\mathbf{r} \cdot \mathbf{R})^2 = \frac{d}{dt} \mathbf{r}(\mathbf{r} \cdot \mathbf{R})^2 - 2\mathbf{r}(\mathbf{v} \cdot \mathbf{R})(\mathbf{r} \cdot \mathbf{R})$$

---

[5] Only the part of the force dependent upon the vector potential is calculated.

and

$$\mathbf{v}r^2 = \frac{d}{dt}\mathbf{r} \times r^2 - 2\mathbf{r}\,(\mathbf{r}\cdot\mathbf{v})$$

The total time derivatives, which vanish in the averaging process, are discarded in advance. Transform the left-hand sides of the equations as follows:

$$\mathbf{v}\,(\mathbf{r}\cdot\mathbf{R})^2 = \frac{2}{3}\mathbf{v}\,(\mathbf{r}\cdot\mathbf{R})^2 - \frac{2}{3}\mathbf{r}\,(\mathbf{v}\cdot\mathbf{R})\,(\mathbf{r}\cdot\mathbf{R}) = \frac{2}{3}\,[(\mathbf{r}\times\mathbf{v})\times(\mathbf{r}\cdot\mathbf{R})\,\mathbf{R}]$$

$$\mathbf{v}r^2 = -\frac{2}{3}\,[\mathbf{r}\times(\mathbf{r}\times\mathbf{v})]$$

As a result the quadrupole term of the vector potential reduces to the form

$$\mathbf{A}_q = \frac{1}{3c}\sum_i \overline{e_i\,[(\mathbf{r}^i\times\mathbf{v}^i)\times(\mathbf{r}^i\cdot\nabla)\,\mathrm{grad}\,(1/R)]}$$

Rewriting the obtained equation in tensor notation, we find that the quadrupole vector potential is defined by the following tensor:

$$Q_{\beta\lambda} \equiv \frac{1}{3c}\sum_i \overline{(\mathbf{r}^i\times\mathbf{v}^i)_\beta\,x^i_\lambda}, \qquad Q_{\beta\beta} = 0$$

$$(A_q)_\alpha = \varepsilon_{\alpha\beta\gamma}Q_{\beta\lambda}\,\frac{(\delta_{\gamma\lambda}R^2 - 3X_\gamma X_\lambda)}{R^5}$$

2. Study the motion of a magnetic moment $\boldsymbol{\mu}$ in a magnetic field given by the components $H_z = -H_0$, $H_x = H_1\cos\omega t$, $H_y = H_1\sin\omega t$. Consider the cases $\omega = eH_0/(2mc)$ and $\omega \to 0$.

Solution. From (17.31b) we obtain the general precession equation:

$$\frac{d\boldsymbol{\mu}}{dt} = \frac{e}{2mc}\boldsymbol{\mu}\times\mathbf{H}$$

By multiplying both sides of this equation by $\boldsymbol{\mu}$ scalarly, we see that μ^2 is conserved. It is, therefore, sufficient to write the equations only for the components μ_x and μ_y, replacing then μ_z by $(\mu^2 - \mu_x^2 - \mu_y^2)^{1/2}$.

Using the abbreviated notation $\omega_0 = eH_0/(2mc)$ and $\omega_1 = eH_1/(2mc)$, we multiply the equation for μ_y by $\pm i$ and combine with the equation for μ_x to get

$$\frac{d}{dt}\,(\mu_x \pm i\mu_y) = \pm\,i\omega_0\,(\mu_x \pm i\mu_y) \pm i\omega_1 e^{\pm i\omega t}\,(\mu^2 - \mu_x^2 - \mu_y^2)^{1/2}$$

We seek the solution in the form $\mu_x \pm i\mu_y = A_\pm e^{\pm i\omega t}$, and get the following equation for the amplitudes $A_\pm$:

$$(\omega - \omega_0)\,A_\pm = \omega_1\,(\mu^2 - A_+A_-)^{1/2}$$

Multiplying the equation with A_+ in the left-hand side by the equation for A_-, we obtain

$$A_+A_- = \frac{\mu^2\omega_1^2}{(\omega - \omega_0)^2 + \omega_1^2}, \qquad |A_\pm| = \frac{\mu\omega_1}{[(\omega - \omega_0)^2 + \omega_1^2]^{1/2}}$$

When $\omega = \omega_0$ (*paramagnetic resonance*), the moment rotates in the x, y-plane with frequency ω_0. When $\omega \to 0$, that is, in the case of an infinitely slow rotation of the field, the moment strictly follows the field.

18

PLANE ELECTROMAGNETIC WAVES

Solution of the Wave Equation. All the results of electrostatics and magnetostatics can be obtained without the help of Maxwell's equations, on the basis of the fundamental laws of Coulomb and Biot-Savart. The essential innovation which could not have preceded Maxwellian physics is the concept of electromagnetic waves. They propagate in vacuum, in the absence of charges, and are obtained as special solutions of Maxwell's equations. These solutions not only led to the conclusion concerning the electromagnetic nature of light. They were the basis for predicting the existence of other electromagnetic waves, both shorter and longer than light waves, notably radio waves.

In the absence of charges or currents, Eqs. (12.43) and (12.44) for scalar and vector potentials are written thus:

$$\nabla^2 \mathbf{A} - \frac{1}{c^2} \frac{\partial^2 \mathbf{A}}{\partial t^2} = 0 \qquad (18.1)$$

$$\nabla^2 \varphi - \frac{1}{c^2} \frac{\partial^2 \varphi}{\partial t^2} = 0 \qquad (18.2)$$

with the additional Lorentz condition (12.42):

$$\operatorname{div} \mathbf{A} + \frac{1}{c} \frac{\partial \varphi}{\partial t} = 0 \qquad (18.3)$$

Equations (18.1) and (18.2) are called *wave equations*. Such equations do not have static solutions at all. Indeed, if the second time derivatives of the potentials are put equal to zero, what remains is $\nabla^2 \mathbf{A} = 0$, $\nabla^2 \varphi = 0$. But if these equations are satisfied over all space, their solutions can only be constant quantities. This is seen from Exercise 1, Section 16.

Let a certain function f satisfy the equation $\nabla^2 f = 0$ over all space and never tend to infinity. There is no point at which it can have a maximum. Assuming the reverse, the maximum point could be surrounded by a small sphere on which the function assumes only smaller values than at the centre, which contradicts the result of the exercise mentioned above. Consequently, nowhere does function f,

or any of its derivatives, which satisfy the same equation, have a maximum, and they are everywhere finite. Only a constant quantity can possess such a property.

Thus, wave equations have only nonstationary solutions valid over all space. We shall look for such partial solutions of (18.1) and (18.2) which depend on one Cartesian coordinate, for example x, and on time. Such solutions are, apparently, subject to the following equations:

$$\frac{\partial^2 \mathbf{A}}{\partial x^2} - \frac{1}{c^2} \frac{\partial^2 \mathbf{A}}{\partial t^2} = 0 \tag{18.4}$$

$$\frac{\partial^2 \varphi}{\partial x^2} - \frac{1}{c^2} \frac{\partial^2 \varphi}{\partial t^2} = 0 \tag{18.5}$$

and the supplementary condition

$$\frac{\partial A_x}{\partial x} + \frac{1}{c} \frac{\partial \varphi}{\partial t} = 0 \tag{18.6}$$

We shall find the most general solution of this set of equations. We temporarily introduce the following notation:

$$x + ct = \xi, \qquad x - ct = \eta \tag{18.7}$$

We transform (18.5) to these independent variables ξ and η. Equation (18.4) can be rewritten symbolically as

$$\left(\frac{\partial}{\partial x} + \frac{1}{c} \frac{\partial}{\partial t} \right) \left(\frac{\partial}{\partial x} - \frac{1}{c} \frac{\partial}{\partial t} \right) \varphi = 0 \tag{18.8}$$

Then

$$\frac{\partial \varphi}{\partial x} = \frac{\partial \varphi}{\partial \xi} \cdot \frac{\partial \xi}{\partial x} + \frac{\partial \varphi}{\partial \eta} \frac{\partial \eta}{\partial x} = \frac{\partial \varphi}{\partial \xi} + \frac{\partial \varphi}{\partial \eta}$$

$$\frac{1}{c} \frac{\partial \varphi}{\partial t} = \frac{\partial \varphi}{\partial \xi} \frac{1}{c} \frac{\partial \xi}{\partial t} + \frac{\partial \varphi}{\partial \eta} \frac{1}{c} \frac{\partial \eta}{\partial t} = \frac{\partial \varphi}{\partial \xi} - \frac{\partial \varphi}{\partial \eta}$$

because, for constant t ($dt = 0$), $\partial \xi / \partial x = 1$ and $\partial \eta / \partial x = 1$, while for constant x ($dx = 0$), $(1/c)(\partial \xi / \partial t) = - (1/c)(\partial \eta / \partial t) = 1$ in accordance with the same equations. Thus symbolically

$$\frac{\partial}{\partial x} + \frac{1}{c} \frac{\partial}{\partial t} = 2 \frac{\partial}{\partial \xi}, \qquad \frac{\partial}{\partial x} - \frac{1}{c} \frac{\partial}{\partial t} = 2 \frac{\partial}{\partial \eta}$$

so that in terms of the variables ξ and η Eqs. (18.4) and (18.5) are written in the form

$$\frac{\partial^2 \mathbf{A}}{\partial \xi \, \partial \eta} = 0, \qquad \frac{\partial^2 \varphi}{\partial \xi \, \partial \eta} = 0 \tag{18.9}$$

Integrating any of them with respect to ξ, we obtain

$$\frac{\partial \mathbf{A}}{\partial \eta} = \mathbf{C}(\eta), \qquad \frac{\partial \varphi}{\partial \eta} = C'(\eta) \tag{18.10a}$$

It is not difficult now to integrate with respect to η:

$$\mathbf{A} = \int^{\eta} \mathbf{C}(\eta)\,d\eta + \mathbf{C}_1(\xi), \qquad \varphi = \int^{\eta} C'(\eta)\,d\eta + C'_1(\xi) \quad (18.10b)$$

But the integrals of the arbitrary functions $\mathbf{C}(\eta)$ and $C'(\eta)$ with respect to η are essentially new arbitrary functions of that same variable η, so that finally the required solutions in terms of the variables ξ and η have the form

$$\mathbf{A} = \mathbf{A}_1(\eta) + \mathbf{A}_2(\xi), \qquad \varphi = \varphi_1(\eta) + \varphi_2(\xi) \quad (18.11)$$

It is immediately apparent that substitution of such solutions into (18.9) identically yields zeros.

Returning to the variables x and t, we can rewrite the obtained solution as

$$\mathbf{A} = \mathbf{A}_1(x - ct) + \mathbf{A}_2(x + ct)$$
$$\varphi = \varphi_1(x - ct) + \varphi_2(x + ct) \quad (18.12)$$

It contains two arbitrary functions for each of the Eqs. (18.4) and (18.5), so that it is a general solution.

Travelling Plane Waves. The solution depending on $x - ct$ is not connected with the solution whose argument is $x + ct$; these are two linearly independent solutions. It is therefore sufficient to investigate one of them:

$$\mathbf{A} = \mathbf{A}(x - ct) \quad (18.13)$$

$$\varphi = \varphi(x - ct) \quad (18.14)$$

In order to satisfy the supplementary condition (18.6), we perform a gauge transformation:

$$\varphi(x - ct) = \varphi'(x - ct) - \frac{1}{c}\frac{\partial}{\partial t}f(x - ct) = \varphi' + \dot{f} \quad (18.15)$$

(the dot over f denotes differentiation with respect to the argument $\eta = x - ct$). But if we put $\varphi' = -\dot{f}$, we obtain simply $\varphi = 0$. Then from (18.6) we also obtain $A_x = 0$. Thus, for a solution of the form considered, depending on $x - ct$ only, the Lorentz condition is satisfied most simply by substituting $\varphi = 0$, $A_x = 0$.

The electric field component along x is equal to zero:

$$E_x = -\frac{1}{c}\frac{\partial A_x}{\partial t} - \frac{\partial \varphi}{\partial x} = 0 \quad (18.16)$$

Since the field does not depend on the gauge transformation of the potentials, the result is a general one.

The magnetic field component along x is also equal to zero:

$$H_x = \frac{\partial A_z}{\partial y} - \frac{\partial A_y}{\partial z} = 0 \qquad (18.17)$$

We find the remaining field components:

$$E_y = -\frac{1}{c}\frac{\partial A_y}{\partial t} = \dot{A}_y, \qquad E_z = -\frac{1}{c}\frac{\partial A_z}{\partial t} = \dot{A}_z$$

$$H_y = -\frac{\partial A_z}{\partial x} = -\dot{A}_z, \qquad H_z = \frac{\partial A_y}{\partial x} = \dot{A}_y \qquad (18.18)$$

From this it follows that **E** and **H** are perpendicular, because

$$\mathbf{EH} = E_y H_y + E_z H_z = 0 \qquad (18.19)$$

They are equal in absolute magnitude, since $|\mathbf{E}| = |\mathbf{H}| = (|\dot{A}_y|^2 + |\dot{A}_z|^2)^{1/2}$.

The equality and perpendicularity of the electric and magnetic fields are essentially invariant properties of the obtained solution. Indeed, from Eqs. (15.3) and (15.4), $|\mathbf{H}|^2 - |\mathbf{E}|^2$ and **EH** are relativistically invariant quantities. If $|\mathbf{E}| = |\mathbf{H}|$ and $\mathbf{EH} = 0$ in one reference frame, the same equations hold in any reference frame. The solution of the form (18.13) has a simple physical meaning.

Let us take the value of **E** at an instant $t = 0$ on the plane $x = 0$. It is equal to **E** (0). It is clear that **E** (0) will have the same value at the instant of time t on the plane $x = ct$, because $\mathbf{E}(x - ct) = \mathbf{E}(0)$ on that plane. We can also say that the plane on which the field **E** is equal to **E** (0) is translated in space through a distance ct in the time t, that is, it moves with a velocity c. The same applies to any plane $x = x_0$, for which there was some value of field $\mathbf{E}(x_0)$ at the initial instant of time. To summarize, all planes with the given value of field propagate through space with the velocity c. Therefore, the solution $\mathbf{E}(x - ct)$ is called a *travelling plane wave*.

We note that the form of the wave does not change as it moves; the distance between planes $x = x_1$ and $x = x_2$, for which **E** is equal to $\mathbf{E}(x_1)$ and $\mathbf{E}(x_2)$, is constant. This result holds for any arbitrary form of wave travelling in free space.

Repeating, the velocity of propagation of a wave in vacuum does not depend on its shape or amplitude and is equal to the universal constant c.

The electric and magnetic fields, as we have seen from (18.19), are perpendicular to the direction of wave propagation, as well as to each other. This is why it is said that electromagnetic waves are transverse (as opposed to longitudinal sound waves in air, for which the oscillations occur in the direction of propagation).

If we denote a unit vector in the direction of the propagation of the wave by **n** and plot it along the x axis of a right-handed coordi-

nate system, the electric field will be directed along the y axis, and the magnetic along the z axis. These directions correspond to the thumb, index and middle fingers of the right hand. We will always describe electromagnetic waves using a coordinate system in which the electric field coincides with the y axis.

Since the vector **n** is directed along the x axis, we can write $x = $ $= $ **rn,** and the electric field can be represented in the form

$$E_y = \dot{A}_y \,(\mathbf{rn} - ct) \tag{18.20}$$

But in this notation it is no longer necessary to relate vector **n** to the x axis. A solution with an argument of the same form as in (18.20) is applicable to any direction of **n**, provided only that **n**, **E** and **H** are mutually perpendicular.

We shall now find the mechanical integrals of motion for an electromagnetic wave.

According to (15.25), the momentum density of an electromagnetic wave is

$$\frac{1}{4\pi c} \, (\mathbf{E} \times \mathbf{H})_x = \frac{1}{4\pi c} \, \dot{A}_y^2 \tag{18.21}$$

and the energy density, from (15.24d), is

$$\frac{|\,\mathbf{E}\,|^2 + |\,\mathbf{H}\,|^2}{8\pi} = \frac{1}{4\pi} \, \dot{A}_y^2 \tag{18.22}$$

It differs from the momentum density by the factor c. That is precisely the property of the energy and momentum of a particle of zero mass, as in Eq. (14.13). This circumstance is extremely important for the quantum theory of light (see Part III).

The density of the energy flux of an electromagnetic field is expressed by Eq. (15.26). From it, the Poynting vector of a plane electromagnetic wave is equal to

$$\frac{c}{4\pi} \, \mathbf{E} \times \mathbf{H} = \frac{c\mathbf{n}}{4\pi} \, \dot{A}_y^2 \tag{18.23}$$

which agrees with the energy-density expression (18.22).

The spatial part of the energy-momentum tensor of an electromagnetic field can be used to calculate the pressure exerted by a plane electromagnetic wave. From the general formula (15.27) it is apparent that, when only E_y and H_z are other than zero, there remains one component T_{xx}, which is equal to

$$\frac{1}{8\pi} \, (E_y^2 + E_z^2 - E_x^2 + H_y^2 + H_z^2 - H_x^2) = \frac{1}{4\pi} \, \dot{A}_y^2 \tag{18.24}$$

This component represents the momentum along the x axis crossing a unit area perpendicular to the axis in unit time. If an incident wave is normal to an absorbing barrier, the whole momentum is

transferred to the barrier. But according to Newton's Second Law (see (1.1)), the momentum transported in unit time equals the force exerted on a unit area of a barrier normal to it. Hence, T_{xx} represents the pressure exerted by an incident normal electromagnetic wave on an absorbing barrier. We find that the pressure of a plane wave is equal to the density of its energy.

This prediction of electrodynamics was experimentally confirmed by P. N. Lebedev. It was demonstrated that an electromagnetic field can indeed be treated as a mechanical system, as is done in Section 15 of this book.

Harmonic Waves. Special interest is attached to travelling waves for which the function $\mathbf{E}\,(x - ct)$ is harmonic. The most general harmonic solution is of the following form:

$$\mathbf{E} = \mathrm{Re}\,[\mathbf{F}e^{-i\omega(t - r\mathbf{n}/c)}] \tag{18.25}$$

where the symbol Re denotes the real part of the expression inside the brackets, $\mathbf{F}$ is a complex vector of the form $\mathbf{F_1} + i\mathbf{F_2}$ (cf. (7.14c)), and ω is the wave frequency in the same sense as in (7.3); ω is the number of radians per second by which the argument of the exponential changes.

The vector $\omega\mathbf{n}/c$ is called the *wave vector*. It is denoted by $\mathbf{k}$:

$$\mathbf{k} \equiv \frac{\omega}{c}\,\mathbf{n}, \qquad |\mathbf{k}|^2 = \frac{\omega^2}{c^2} \tag{18.26}$$

The geometric meaning of $\mathbf{k}$ is easy to explain. We define the wavelength, that is, the spatial distance Δr over which $\mathbf{E}$ reverts to the same value. Let the required wavelength be λ. Then

$$e^{i\omega\lambda/c} = e^{i\omega(\mathbf{n}\cdot\Delta\mathbf{r})/c} = e^{i|\Delta\mathbf{r}||\mathbf{k}|} = e^{2\pi i} \tag{18.27}$$

because the period of the function e^{ix} is equal to 2π. Hence

$$\lambda = \frac{2\pi c}{\omega} \tag{18.28}$$

Comparing the wavelength with the wave vector, we obtain

$$\mathbf{k} = \frac{2\pi}{\lambda}\,\mathbf{n}, \qquad \lambda = \frac{2\pi}{|\mathbf{k}|} \tag{18.29}$$

Sometimes a quantity smaller than the wavelength by a factor of 2π, and denoted $\bar{\lambda}$, is used. It is equal to the inverse value of the absolute magnitude of the wave vector.

Let us see how the frequency and wave vector of a harmonic electromagnetic wave transform in passing from one inertial frame of reference to another. We shall show that the transformation properties of the components of the wave vector and the frequency are the same as those of the coordinates and time.

We shall introduce the concept of *wave phase* as an argument of the exponential function (18.25) and prove that phase is an invariant quantity. Indeed, phase characterizes a certain event, say the vanishing of an electric and magnetic field at some instant of time at some point in space. If that same wave is considered in another reference frame, the coordinates and time corresponding to that event will have other values, but the event itself, the fact, will not, of course, have changed.

This is easily understood by imagining an electric field being measured according to the readings of some noninertial instrument. Two such instruments superimposed at some instant at the same point in space but having a relative velocity of motion must show zero reading for the field, otherwise the reference frame in which the electromagnetic field is zero will be in some way preferred in comparison with the others. For example, a light-sensitive plate exposed at a given instant would fail to darken in that frame of reference.

According to (18.25), the expression for the phase of a wave is

$$\psi = \mathbf{kr} - \omega t = x_1 k_x + x_2 k_y + x_3 k_z + x_4 \frac{i\omega}{c} \qquad (18.30)$$

For it to be invariant the wave vector and frequency must be assumed to constitute together a single four-vector: $k_x = k_1$, $k_y = k_2$, $k_z = k_3$, $i\omega/c = k_4$. Then the phase is written as $\psi = k_i x_i$.

The four-vector k_i possesses a peculiar property, which is apparent from the second, scalar equation (18.26):

$$|\mathbf{k}|^2 - \frac{\omega^2}{c^2} = k_i k_i = k_i^2 = 0 \qquad (18.31)$$

In other words, it is said that k_i is a zero vector, that is, a vector of zero four-dimensional length. Unlike three-vectors, which can have zero length only when all three components vanish, four-vectors do not possess this property.

Applying the general formulas (13.33) and (13.34) for the transformation of the components of a four-vector, we obtain

$$k_x = \frac{k_x' + \omega' V/c^2}{(1 - V^2/c^2)^{1/2}}, \qquad k_x' = \frac{k_x - \omega V/c^2}{(1 - V^2/c^2)^{1/2}} \qquad (18.32)$$

$$\omega = \frac{\omega' + V k_x'}{(1 - V^2/c^2)^{1/2}}, \qquad \omega' = \frac{\omega - V k_x}{(1 - V^2/c^2)^{1/2}} \qquad (18.33)$$

As usual, the projections k_y and k_z, which are perpendicular to the relative velocity of the reference frames, do not change.

Suppose now that a light source is at rest relative to the primed reference frame, that is, it is moving with velocity V along the x axis of the unprimed, "stationary", system. A stationary observer

measures the frequency of a light beam travelling at an angle Θ to the x axis. Then, from (18.33) and (18.26), and substituting $k_x = (\omega/c) \cos \Theta$, we find the change in the frequency of the moving light source as compared with its frequency in its proper reference frame:

$$\omega' = \frac{\omega \left[1 - (V/c) \cos \Theta\right]}{(1 - V^2/c^2)^{1/2}} \tag{18.34}$$

If $\Theta < \pi/2$, then $\omega > \omega'$.

This formula describes the well-known *Doppler effect*, which is used to measure the ray velocities (that is, directed strictly along the line of vision) of celestial bodies. To the accuracy of the term linear in V/c, the effect is obtained in nonrelativistic theory from elementary kinematic considerations.

The significance of the relativistic formula is especially apparent when the source is moving perpendicular to the line of vision. Then, in the second equation in (18.33) we must put $k_x = 0$, obtaining thus

$$\omega = \omega' (1 - V^2/c^2)^{1/2} \tag{18.35}$$

This formula expresses the *transverse Doppler effect*. It is directly associated with the relativistic time dilation described by Eq. (13.21a). A moving emitter of harmonic waves can be treated as a clock, insofar as a periodic process with a frequency ω' takes place in it. The total number of oscillations is an invariant quantity: every oscillation is an event. Since a moving clock shows less time to have passed, the oscillation frequency must be correspondingly greater.

The transverse Doppler effect has been observed spectroscopically with atoms in motion (Ives-Stilwell experiment, 1938). For this experiment the ratio V/c was sufficient to observe the frequency shift. This offered direct experimental proof of time dilation in relative motion.

Polarization of a Plane Harmonic Wave. Let us now study the nature of the oscillations of an electric field in a plane harmonic (otherwise called "monochromatic") wave. For this, we write the vector **F** (see (18.25)) in the form

$$\mathbf{F} = \mathbf{F}_1 + i\mathbf{F}_2 = (\mathbf{E}_1 - i\mathbf{E}_2) e^{i\alpha} \tag{18.36}$$

We choose the phase α so that the vectors $\mathbf{E}_1$ and $\mathbf{E}_2$ are mutually perpendicular. We multiply Eq. (18.36) by $e^{-i\alpha}$ and square. Then we obtain

$$(\mathbf{E}_1 - i\mathbf{E}_2)^2 = |\,\mathbf{E}_1\,|^2 - |\,\mathbf{E}_2\,|^2$$
$$= e^{-2i\alpha} [\,|\,\mathbf{F}_1\,|^2 - |\,\mathbf{F}_2\,|^2 + 2i\,(\mathbf{F}_1 \cdot \mathbf{F}_2)] \tag{18.37}$$

We have taken advantage of the fact that $\mathbf{E}_1$ and $\mathbf{E}_2$ are perpendicular. Because of this $(\mathbf{E}_1 - i\mathbf{E}_2)^2$ is a purely real quantity. There-

fore, the imaginary part of (18.37) must be put equal to zero. Representing $e^{-2i\alpha}$ as $\cos 2\alpha - i \sin 2\alpha$, we obtain

$$- (\mid F_1 \mid^2 - \mid F_2 \mid^2) \sin 2\alpha + 2 (F_1 \cdot F_2) \cos 2\alpha = 0$$

or

$$\tan 2\alpha = \frac{2 (F_1 \cdot F_2)}{\mid F_1 \mid^2 - \mid F_2 \mid^2} \tag{18.38}$$

whence the angle α is determined for the given solution (18.25).

It is now easy to express E_1 and E_2. By (18.36), $E_1 - iE_2 = = (F_1 + iF_2) e^{-i\alpha} = F_1 \cos \alpha + F_2 \sin \alpha - i (F_1 \sin \alpha - F_2 \cos \alpha)$, so that

$$E_1 = F_1 \cos \alpha + F_2 \sin \alpha, \qquad E_2 = F_1 \sin \alpha - F_2 \cos \alpha \tag{18.39}$$

We now include the constant phase α in the exponent of (18.25) and, for short, put

$$\alpha - \omega (t - rn/c) \equiv \psi \tag{18.40}$$

Then, in the most general case, the electric field for a plane harmonic wave will be

$$E = \mathrm{Re}\,[(E_1 - iE_2)\, e^{i\psi}] = E_1 \cos \psi + E_2 \sin \psi \tag{18.41}$$

Here, the vectors E_1 and E_2 are defined as perpendicular.

Let us now represent the solution of (18.41) graphically. Let the wave propagate along the x axis. The y axis is directed along E_1, and z axis along E_2. Hence, from (18.41) we obtain

$$E_y = \mid E_1 \mid \cos \psi, \qquad E_z = \mid E_2 \mid \sin \psi \tag{18.42}$$

We eliminate the phase ψ. For this divide the first equation by $\mid E_1 \mid$, the second by $\mid E_2 \mid$, square and add. Then the phase is eliminated and an equation relating the field components remains:

$$\frac{E_y^2}{\mid E_1 \mid^2} + \frac{E_z^2}{\mid E_2 \mid^2} = 1 \tag{18.43}$$

It follows that the electric field vector describes an ellipse in the y,z-plane, which is itself moving along the x axis with velocity c, and passes around the whole ellipse over one wavelength. Relative to a fixed coordinate system, the electric field vector describes a helix wound on an elliptic cylinder. The pitch of the helix is equal to the wavelength.

Such an electromagnetic wave is termed *elliptically polarized*. It represents the most general form of a plane harmonic wave (18.25).

If one of the components is equal to zero, for example $E_1 = 0$ or $E_2 = 0$, then the oscillations of E occur in one plane. Such a wave is termed *plane polarized*.

When $| \mathbf{E}_1 |$ is equal to $| \mathbf{E}_2 |$, the vector $\mathbf{E}$ describes a circle in the y,z-plane. Depending on the sign of $\mathbf{E}_2$, the rotation around the circle occurs in a clockwise or conterclockwise direction. Accordingly, the wave is termed *right-hand* or *left-hand polarized*. Figure 25 shows the configuration of vectors $\mathbf{E}_1$, $\mathbf{E}_2$, and $\mathbf{E}$ for a right-hand polarized wave and a left-hand polarized wave. For the same value of phase ψ, the rotation is either clockwise or counterclockwise.
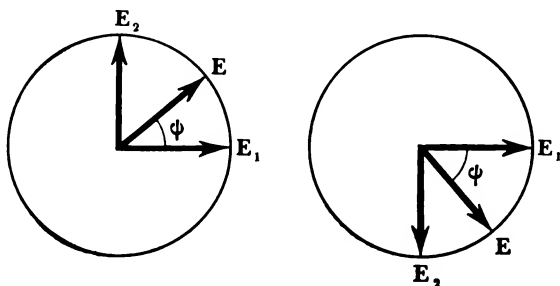
Figure 25

The sum of two circularly polarized waves of equal amplitude gives a plane polarized wave. The relationship between their phases determines the plane of polarization. Thus, if the waves shown in Figure 25 are added the oscillations $\mathbf{E}_2$ and $-\mathbf{E}_2$ mutually cancel out, and only the plane polarized oscillation $\mathbf{E}$ remains. In turn, a circularly polarized oscillation is resolved into two mutually perpendicular plane oscillations.

In nature it is most common to observe unpolarized (natural) light. Naturally, such light cannot be strictly monochromatic (that is, possessing strictly one frequency ω), for, as we have just shown, monochromatic light is always polarized in some way. But if we imagine that the components $\mathbf{E}_1$ and $\mathbf{E}_2$ in Figure 25 are not related by a strict phase relationship (18.42) but randomly change their relative phases, then the resultant vector will also change its direction in the y,z-plane in a random manner. For this, it is necessary that the oscillation frequencies should vary within some finite interval $\Delta\omega$, since the difference of phase between two oscillations of strictly constant and identical frequency is constant.

Natural light scattered on charges may be polarized for certain directions of the scattered beam.

EXERCISES

1. Show that the superposition of two waves of equal amplitude, circularly polarized in opposite directions and travelling in the same direction, the wavelength difference between them being $\Delta\lambda$, produces a plane polarized wave whose polarization vector rotates as the wave propagates.

2. Find the relationship between angles θ and θ' defining the inclination of a light beam to the direction of the relative velocity of two reference frames.

Solution. Substituting $k_x = (\omega/c)\cos\theta$, $k'_x = (\omega'/c)\cos\theta$ into the first equations of (18.32), and taking into account that $k_y = (\omega/c)\sin\theta$, $k'_y = (\omega'/c)\sin\theta'$, $k_y = k'_y$, we find

$$\tan\theta = \frac{\sin\theta'}{\cos\theta' + V/c}\,(1 - V^2/c^2)^{1/2}$$

whence

$$\cos\theta = \frac{\cos\theta' + V/c}{1 + (V/c)\sin\theta'}$$

which agrees with the velocity addition formula (13.23).

In applying the obtained relationship to the phenomenon of light aberration, we must put $\theta' = \pi/2$ for the rays of a star passing perpendicular to the plane of the earth's orbit. Then in a reference frame fixed relative to the earth we obtain a constant angle of inclination of the ray to the perpendicular to the plane of the orbit:

$$\frac{\pi}{2} - \theta \approx \frac{V}{c}$$

In the earth's annual motion the respective star describes a circle of angular radius V/c. Taking into account that the star is at a finite distance from the sun, its parallax is additionally superimposed on this angular displacement.

3. Prove that the quantity $\omega^2\,d\Omega$ is relativistically invariant.

Hint. Make use of Eq. (18.34) and the result of Exercise 2.

19

TRANSMISSION OF SIGNALS. ALMOST PLANE WAVES

The Impossibility of Transmitting a Signal by Means of a Monochromatic Wave. A plane monochromatic wave (18.25) extends without limit in all directions of space and in time. Nowhere, so to speak,

does it have a beginning or an end. What is more, its properties are everywhere always the same: its frequency, amplitude and the distance between two travelling crests (that is, the wavelength λ) are always constant. All this can be easily seen by considering a sinusoid or helix.

Let us now pose the problem of the possibility of transmitting an electromagnetic signal over a distance. In order to transmit the signal, an electromagnetic disturbance must be concentrated in a certain volume. By propagation, this disturbance can reach another region of space; detected by some means (for example, a radio receiver), it will transmit to the point of reception a signal about an event occurring at the point of transmission. Likewise, our visual perceptions are a continuous recording of electromagnetic (light) disturbances originating in surrounding objects. A signal must somehow be bounded in time in order to give notice of the beginning and end of any event.

In order to transmit a signal the amplitude of the wave must, for a time, be somehow changed. For example, the amplitude of one of the waves of the sinusoid must be increased and we must wait until this increased amplitude arrives at the receiving device. A strictly monochromatic wave, that is, a sinusoid, has the same amplitude everywhere and is therefore not suitable for transmitting signals. In the same way, an ideal plane wave with a given wave vector cannot transmit the image of an object limited in space.

Propagation of a Nonmonochromatic Wave. Let us now see how the superposition of several sinusoids, that is, monochromatic waves, can be used to transmit signals. Suppose that we have at our disposal a frequency spectrum of travelling waves lying within the interval $\omega_0 - \Delta\omega/2 \leqslant \omega \leqslant \omega_0 + \Delta\omega/2$, such that the total width of the spectrum, that is, the frequency interval $\Delta\omega$, is considerably smaller than the *carrier frequency* ω_0. For the sake of simplicity, the amplitudes of all the waves will be assumed to be identical and equal to $\mathbf{E}_0(\omega) = \mathbf{E}_0$ within the chosen spectrum, vanishing outside that interval.

Then the resulting oscillation will be represented by the integral over all the separate monochromatic oscillations:

$$\mathbf{E} = \int \mathbf{E}_0(\omega)\, e^{-i(\omega t - kx)}\, d\omega$$

$$= \mathbf{E}_0 \int_{\omega_0 - \Delta\omega/2}^{\omega_0 + \Delta\omega/2} e^{-i(\omega t - kx)}\, d\omega \tag{19.1}$$

In this equation not only the frequency is variable, but also the absolute value of the wave vector $\mathbf{k}$, the so-called *wave number*.

According to (18.26) it is equal to ω/c, but here it is more convenient to consider a dependence of more general form: $k = k(\omega)$.

Since the frequency lies within a small interval, k can be expanded in a power series in $(\omega - \omega_0)$:

$$k(\omega) = k(\omega_0) + (\omega - \omega_0)\left(\frac{dk}{d\omega}\right)_0 \tag{19.2}$$

Substituting into (19.1), we obtain the following expression for the field:

$$\mathbf{E} = \mathbf{E}_0 e^{-i(\omega_0 t - k_0 x)} \int_{\omega_0 - \Delta\omega/2}^{\omega_0 + \Delta\omega/2} e^{-i(\omega - \omega_0)[t - (dk/d\omega)_0 x]}\, d\omega \tag{19.3}$$

We now introduce a new integration variable $\xi = \omega - \omega_0$. Evaluating the integral, we reduce it to the form

$$\mathbf{E} = \mathbf{E}_0 e^{-i(\omega_0 t - k_0 x)} \int_{-\Delta\omega/2}^{\Delta\omega/2} e^{-i\xi[t - (dk/d\omega)_0 x]}\, d\xi$$

$$= \mathbf{E}_0 e^{-i(\omega_0 t - k_0 x)} \times \frac{2 \sin\{[t - (dk/d\omega)_0\, x]\,(\Delta\omega/2)\}}{t - (dk/d\omega)_0\, x} \tag{19.4}$$

Let us now examine the expression obtained. It consists of two factors. The first of them, $\mathbf{E}_0 e^{-i(\omega_0 t - k_0 x)}$, represents a travelling wave homogeneous in space with a mean carrier frequency ω_0. However, the amplitude of the resultant wave is no longer constant in space because of the second factor

$$\frac{2 \sin\{[t - (dk/d\omega)_0\, x]\,(\Delta\omega/2)\}}{t - (dk/d\omega)_0\, x} \equiv g\left\{[t - (dk/d\omega)_0\, x]\,\frac{\Delta\omega}{2}\right\} \equiv g(\chi)$$

where the designations g and χ are obvious from the equation. Thanks to the sine, this factor has an infinite number of maxima. One of them, however, is the greatest and is attained when the argument of function χ vanishes. The other maxima are smaller and decrease with distance from the principal one, which is located at $x = (d\omega/dk)_0 t$. It can thus be observed that this maximum has no constant position in space and itself moves with a velocity

$$v = \frac{d\omega}{dk} \tag{19.5}$$

because it follows from the definition of the maximum point $\chi = 0$ that

$$x = \frac{d\omega}{dk}\, t = vt$$

As mentioned at the beginning of this section, the displacement of the maximum can be applied for transmitting signals from one

spatial point to another, because this maximum is distinguished from the other maxima. Such a spatially concentrated disturbance is called a *wave packet.*

A wave packet need not necessarily have the form shown in Figure 26, where it refers to the expression (19.4). By selecting a relationship $E_0(\omega)$ other than in Eq. (19.1), that is, one involving not a constant amplitude in the frequency interval $\Delta\omega$ but a more complex function of frequency, the shape of $E(x)$ can be changed. In particular, it is simple to make the resultant amplitude rectangular, so that the transmitted signal resembles a Morse-code dash.
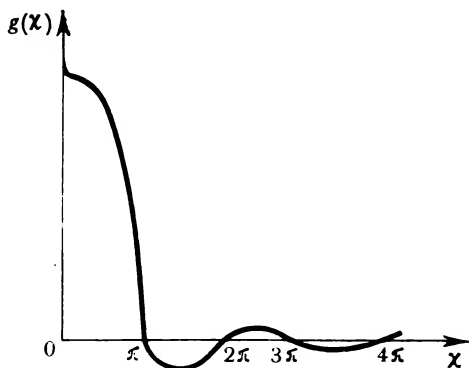
Figure 26

The word "simple" here refers to the analytical definition of the quantity $E_0(\omega)$ yielding a rectangular signal. Indeed, Eq. (19.1) is in fact an integral Fourier transformation from function $E_0(\omega)$ of variable ω to function $g(\chi)$ of variable χ. But the Fourier transformation possesses the property of reciprocity: if function $E_0(\omega)$ is given the form corresponding to $g(\chi)$ in Eq. (19.4), the output signal $E_0(x)$ will be rectangular.

If the carrier frequency ω_0 is high enough, it can be used to transmit separate audio-frequency signals without their superimposing on one another. In other words, it is possible to reproduce music or speech without appreciable distortions.

Frequency Range and Signal Duration. As we have seen, a certain frequency range is always required to transmit a signal. A monochromatic wave with a strictly definite frequency is uniform in time: it, as it were, transmits one signal of infinite duration. From the Fourier inversion theorem, which expresses the property of reciprocity of the integrals, we conclude that a signal of infinitesimal duration requires an infinitely large frequency range for transmission.

But what if the duration of the signal is finite? What frequency range is needed to transmit it?

This can be concluded from an examination of Figure 26, which presents the dependence of g on χ. For the purpose of transmitting a signal, only the domain of the curve close to the principal maximum at $\chi = 0$ is important. In units of χ it is of the magnitude of π. Hence, the duration of the signal is determined from the equation

$$\Delta\chi = \frac{\Delta\omega}{2}\,\Delta t \sim \pi$$

In other words, to transmit a signal of duration Δt the required frequency interval is $\Delta\omega$, connected with Δt by the relationship

$$\Delta\omega\,\Delta t \sim 2\pi \qquad\qquad (19.6a)$$

It should be noted that this estimate refers only to the order of magnitude of $\Delta\omega$ and Δt. Determination of $\Delta\chi$ is to a certain extent arbitrary and done purely visually from the shape of the curve.

The shape of the curve suggests that the main contribution to the transmission of the signal is made by small values of χ, close to zero, because around zero the amplitude of the wave is greater. From an evaluation of the frequency interval according to the relative contribution of the various frequencies to the transmitted signal, it is possible to develop a somewhat more strict evaluation than (19.6a), namely

$$\langle\Delta\omega\rangle\,\langle\Delta t\rangle \sim 1 \qquad\qquad (19.6b)$$

Here the quantities are in angular brackets to stress that the method of evaluation is different than in (19.6a). It must, furthermore, be pointed out that (19.6a) and (19.6b) are lower estimates. In many cases the strong inequality $\Delta\omega\,\Delta t \gg 2\pi$ holds.

If a radio station is required to transmit sounds audible to the human ear, then the quantity Δt must not be greater than 0.5×10^{-4} s, since the limit of audibility is 2×10^4 oscillations per second. Actually, frequencies not higher than 0.5×10^4 are adequate for transmission.

The frequency range $\Delta\omega$ is always less than the carrier frequency ω_0, which, even for the longest-wave transmitting stations, is not less than 10^6. The frequency ω_0 must be compared with an interval $\Delta\omega$ of the order of 0.5×10^4, since cutting off the highest frequencies in music, singing or speech does not introduce any essential distortion.

Television transmissions require a considerably greater frequency interval, because an image must be reproduced 25 times every second and, in turn, consists of tens of thousands of separate points. As a result, the carrier frequency must be very high, corresponding

to the metre waveband. Such waves propagate within a relatively small radius; they are screened by the curvature of the earth's surface, just as light is. It is interesting that the transmission of colour images requires practically the same frequency range as black-and-white images. This is mainly due to the fact that man's colour vision is not as sharp as his contour vision.

Phase and Group Velocity. Let us consider in greater detail the speed with which signals are transmitted. If we apply formula (19.5) to the propagation of a signal in vacuum, we obviously obtain $v = c$. The situation is different in a nonabsorbing material medium, where the dependence of k on ω should be assumed different than in vacuum (the situation is greatly complicated in an absorbing medium). The phenomenon is known as *dispersion of electromagnetic waves.*

We shall not investigate it here and simply accept that between k and ω there is a dependence which is not a direct proportionality. This is extremely important for the optico-mechanical analogy with which Section 21 deals (see Part III). Thus, we take the velocity of a wave packet as being

$$v = \frac{d\omega}{dk}$$

It differs from the propagation velocity of the constant phase surface, which is expressed in terms of frequency and wave number as

$$u = \frac{\omega}{k} \tag{19.7}$$

Indeed, the expression for a travelling monochromatic wave can be written in the following form:

$$E = E_0 e^{ik(x - \omega t/k)}$$

Comparing this formula with the general expression for a travelling wave $E = E(x - ut)$, we arrive at (19.7). The velocity of the wave is designated u rather than c, because (19.7) refers to the propagation of the wave not in vacuum, where the difference between u and c vanishes, but in a nonabsorbing medium; u is the *phase velocity of the wave* and v the *group velocity of the wave packet* resulting from the superposition of a group of waves.

The group velocity can also be defined in vector form:

$$\mathbf{v} = \frac{\partial \omega}{\partial \mathbf{k}} \tag{19.8}$$

thereby defining the direction of the signal.

The Form of a Wave in Space and the Range of the Wave Vectors. An expression similar to (19.6) can also be obtained for the form

of a wave in space at a definite instant of time. For this we **must** take χ at some constant time $t =$ constant and then, once again taking $\Delta\chi \sim \pi$, we obtain

$$\Delta\chi = \frac{\Delta\omega}{2}\frac{dk}{d\omega}\,\Delta x = \frac{\Delta k\,\Delta x}{2} \sim \pi$$

or

$$\Delta k\,\Delta x \sim 2\pi \tag{19.9}$$

For $\langle\Delta k\rangle$ and $\langle\Delta x\rangle$ we can write a formula similar to (19.6b).

In view of the fact that **k** is a vector quantity, relationship (19.9) should be written for all three components of **k**. Then instead of (19.9) we arrive at three estimates:

$$\Delta k_x\,\Delta x \sim 2\pi$$
$$\Delta k_y\,\Delta y \sim 2\pi$$
$$\Delta k_z\,\Delta z \sim 2\pi \tag{19.10a}$$

We shall explain the relations (19.10a) by means of a graphic example. Let us suppose that an electromagnetic wave has, in some way, to be bounded on the sides, as in the case of a radar beam. Let us find the greatest accuracy with which a radar can register the position of an object at a distance l. Obviously, this accuracy is given by the diameter of the beam d at the distance l from the radar.

Let the frequency at which the radar operates be ω, the corresponding wavelength then being $\lambda = 2\pi c/\omega$. If the electromagnetic wave were propagated in unbounded space, it would have (or could have) an accurately defined wave vector

$$\mathbf{k} = \frac{2\pi}{\lambda}\,\mathbf{n}$$

(**n** is the unit vector in the direction of the beam). If the wave has a cross section d, then **k** can no longer be regarded as an accurately defined vector along **n**.

In order to write an expression for the electromagnetic wave at any point in space occupied by the beam, it is necessary to take a group of plane waves whose vectors **k** lie inside a cone described by a certain aperture angle. We shall assume that the axis of the cone coincides with **n**. We have in mind here not a cone with a sharply defined surface but a conic bundle of directions. The dependence of the wave's amplitude on its direction of propagation inside the cone can be, for example, like the curve in Figure 26 close to the principal maximum.

The aperture angle of a conic beam is defined by a certain value of $k_\perp$ in any direction perpendicular to the axis of the cone; $k_\perp$ defines the interval of values, Δk, necessary for the diameter of the bundle

of beams in space to be d. It is apparent that in meaning $k_\perp$ and d represent $\Delta k_y/2$ and Δy, or $\Delta k_z/2$ and Δz in the second and third formulas in (19.10a), if **n** is directed along the x axis. We thus arrive at the relationship

$$2k_\perp d \geqslant 2\pi \qquad\qquad (19.10b)$$

(Δk_y should be put equal to $2k_\perp$, because the divergence from **n** is in both directions).

The dimensions of the radar antenna itself can be ignored if the diameter of the beam is considered at a great distance; and this is
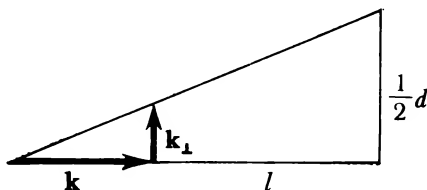
Figure 27

of practical interest. In other words, d is determined only by the relationship (19.10a) and is independent of the dimensions of the antenna.

The divergence of the beam at every point is measured by the ratio $k_\perp/k$. For this reason, the ratio of the cross section of the beam d to the distance from the radar l cannot be less than the quantity $2k_\perp/k$:

$$\frac{d}{l} \geqslant \frac{2k_\perp}{k}$$

This relationship is shown in Figure 27 for the limiting case of the equality, but it should be remembered that what we have here is, rather, the upper estimate of the order of magnitude of $2k_\perp/k$. Equation (19.10b) is the lower estimate of this quantity:

$$\frac{k_\perp}{k} \geqslant \frac{\pi}{dk}$$

Knowing the upper and lower estimates for $k_\perp$, we can eliminate $k_\perp/k$ from them, obtaining

$$\frac{d}{l} \geqslant \frac{2\pi}{dk}, \qquad \text{or} \qquad \frac{d}{l} \geqslant \frac{\lambda}{d}$$

and finally

$$d \geqslant (l\lambda)^{1/2}$$

For example, if $l = 100$ km and $\lambda = 1$ m, then the position of the object cannot be determined with an accuracy exceeding 320 m. This is why the dimensions of the antenna could be neglected.

Limits of Applicability of the Ray Concept. Equations (19.10a) indicate within what limits the concept of a ray is applicable in optics. Obviously, one can talk about a ray in a definite direction only when

$$\Delta k \ll k \qquad\qquad (19.11)$$

that is, when the transverse broadening of the wave vector is considerably less than the wave vector itself. In the radar problem this means that $k_\perp \ll k$. But $k_\perp \sim \pi/d$ and $k = 2\pi/\lambda$, so that (19.11) is equivalent to the condition

$$d \gg \lambda \qquad\qquad (19.12)$$

In other words, the dimensions of the region in which the concept of a light ray is meaningful must be considerably larger than the wavelength of the light wave. For example, a small aperture in the wall of a camera-obscura of diameter, say, 1 mm is considerably greater than the wavelength of visible light, which is of an order of 0.5×10^{-4} cm. Therefore, the image obtained in a camera-obscura is formed with the aid of light rays.

The optics of light rays is called geometrical optics. A ray is defined only when its direction is given, that is, the normal to the wave front. If we are given a beam of nonparallel (for example, converging) rays, then the wave front is curved. But the radius of its curvature at each point must be much greater than the wavelength for it to be represented close to that point with the help of an osculating plane front. Then a convergent beam of rays is represented as the aggregate of normals to the corresponding plane wave fronts.

Close to the focus of the optical system, where all the rays converge, the curvature of the wave front may become comparable with the wavelength, and then deviations from geometrical optics occur. They are known as *wave diffraction*. They are also observed when a wave falls on an opaque obstacle. In accordance with geometrical optics, there should be a sharp shadow—a transition from a domain where the field is not zero to the domain where it is zero. But Maxwell's equations involve derivatives of fields with respect to coordinates and do not permit discontinuous solutions in free space. Actually there is always a transition zone between "light" and "shade" in which the wave amplitude changes in a complex, nonmonotonic way.

EXERCISES

1. Write the relationship between phase and group velocity.

Answer. $v = \dfrac{d\omega}{dk} = \dfrac{duk}{dk} = u + k\,\dfrac{du}{dk} = u - \lambda\,\dfrac{du}{d\lambda}$.

2. Show that if the dependence of the amplitude of a monochromatic wave in a wave packet is proportional to a function of the form

$$\exp\left[(\omega - \omega_0)^2/(\Delta\omega)^2\right]$$

where ω_0 is the carrier frequency, and $\Delta\omega \ll \omega_0$ characterizes the spectrum width, then the form of the wave packet reproduces an analogous function, but of $t - x/u$, a function such that the packet width is inversely proportional to $\Delta\omega$. Compare this with (19.6*a*).

3. Find the limiting size of an object that can be seen under a microscope, using light of wavelength λ.

Solution. Denoting one-half the aperture angle of the cone of the rays from the lens to the object as θ, we have $\Delta k = k \sin\theta$. From this

$$\Delta x \gg \frac{2\pi}{\Delta k} = \frac{2\pi}{k \sin\theta} = \frac{\lambda}{\sin\theta}$$

Thus, it is best to use beams with large aperture angles and short wavelengths.

20

THE EMISSION
OF ELECTROMAGNETIC WAVES

Basic Equations and Boundary Conditions. So far we have considered electromagnetic waves irrespective of the charges producing them. In this section we shall consider the emission of waves by point charges moving in a vacuum. The basic system of equations in this case is (12.43) and (12.44) together with the Lorentz condition (12.42). We rewrite these equations anew:

$$\nabla^2 \mathbf{A} - \frac{1}{c^2}\frac{\partial^2 \mathbf{A}}{\partial t^2} = -\frac{4\pi}{c}\,\mathbf{j} \tag{20.1}$$

$$\nabla^2 \varphi - \frac{1}{c^2}\frac{\partial^2 \varphi}{\partial t^2} = -4\pi\rho \tag{20.2}$$

$$\operatorname{div} \mathbf{A} + \frac{1}{c}\frac{\partial \varphi}{\partial t} = 0 \tag{20.3}$$

Equations (20.1) and (20.2), written for the Cartesian components of the vector potential and for the scalar potential, are of the same form (the Laplacian of a vector differs from the Laplacian of a scalar in curvilinear components, for example A_r, A_θ, A_φ). Therefore we seek the solution of one of these equations and apply it to all the others.

It is convenient to proceed as follows: find the field of one point emitter and then sum over all emitting charges, just as in Section 16 we first obtained the static field of one point charge. By virtue of the linearity of Eqs. (20.1)-(20.3), the solution for an arbitrary distribution of charges and currents is the sum or integral taken over all the emitters.

Suppose that a charge δe, equal to $\rho\, dV$, is placed in a volume element dV at the origin of a coordinate system. Let us look for the corresponding potential.

In order to determine the solution uniquely we must impose a certain boundary condition on it. Let us assume that the charges are located in infinite space free of matter, that is, that there are neither conductors nor dielectrics in the vicinity. Then the boundary condition can be imposed only at an infinite distance from the charges. In accordance with the posed problem, it is natural to assume that there was no radiation field at an infinitely long time before the emission began and at an infinite distance from the emitter:

$$\varphi(t \to -\infty,\ r \to \infty) = 0$$
$$\mathbf{A}(t \to -\infty,\ r \to \infty) = 0 \tag{20.4}$$

This allows for a unique solution of the problem.

Since charge δe is placed at the origin of the coordinate system, it is natural to seek a solution possessing spherical symmetry. Then in the equation for φ we must go over to spherical coordinates and differentiate only with respect to r, as in the electrostatic problem, and also, of course, with respect to time.

We now write Eq. (20.2), leaving only the derivatives with respect to r and t. We assume that it refers to any point except the origin, in which case in the right-hand side we have zero:

$$\frac{1}{r^2}\frac{\partial}{\partial r}\, r^2\, \frac{\partial \varphi}{\partial r} - \frac{1}{c^2}\frac{\partial^2 \varphi}{\partial t^2} = 0 \tag{20.5}$$

Temporarily, we put

$$\varphi = \frac{\Phi(r,\, t)}{r} \tag{20.6}$$

Then

$$\frac{\partial \varphi}{\partial r} = \frac{1}{r}\frac{\partial \Phi}{\partial r} - \frac{\Phi}{r^2}, \qquad r^2\frac{\partial \varphi}{\partial r} = r\frac{\partial \Phi}{\partial r} - \Phi$$

$$\frac{\partial}{\partial r}\, r^2\, \frac{\partial \varphi}{\partial r} = r\frac{\partial^2 \Phi}{\partial r^2} + \frac{\partial \Phi}{\partial r} - \frac{\partial \Phi}{\partial r} = r\frac{\partial^2 \Phi}{\partial r^2}$$

Substituting this into (20.5) and multiplying by r (by definition r is not equal to zero), we obtain

$$\frac{\partial^2 \Phi}{\partial r^2} - \frac{1}{c^2} \frac{\partial^2 \Phi}{\partial t^2} = 0 \tag{20.7}$$

But this is the equation, of the form (18.5), for the propagation of a wave. Its solution is similar to (18.12):

$$\Phi = \Phi_1 \left(t + \frac{r}{c} \right) + \Phi_2 \left(t - \frac{r}{c} \right) \tag{20.8}$$

We now apply the boundary conditions (20.4). The solution Φ_1 depends on the argument $t + r/c$, and the solution Φ_2 depends on the argument $t - r/c$. The first of these arguments, $t + r/c$, for $r \to \infty$ and $t \to -\infty$ has a completely indeterminate form $\infty - \infty$, that is, it is equal to anything. From (20.4) the function Φ becomes zero as $r \to \infty$ and $t \to -\infty$ (r in the denominator of (20.6) is immaterial, as will become apparent from the subsequent discourse). Therefore Φ_1 vanishes for any value of the argument, that is identically. For the function Φ_2 conditions (20.4) denote that $\Phi_2 (-\infty) = 0$. In other words, Φ_2 tends to zero at minus infinity. It does not follow from this, of course, that it is equal to zero everywhere. Thus

$$\Phi = \Phi_2 \left(t - \frac{r}{c} \right)$$

Omitting the 2, we write the expression for φ as follows:

$$\varphi = \frac{1}{r} \Phi \left(t - \frac{r}{c} \right) \tag{20.9}$$

The function Φ is not yet determined. From the form of its argument we conclude that it describes a travelling wave in the direction of increasing values of the radius (because $t > 0$). Such a wave is termed diverging.

Retarded Potential. The value of the function Φ at $r = 0$, $t = 0$ is shifted to the point r in a time $t = r/c$. In other words, the potential at point r and time t is determined by the charge, not at time t, but at an earlier time $t - r/c$. The term r/c is a measure of the retardation occurring as a result of the finite velocity of propagation of the wave. Of course, the change in charge δe located at the origin of the coordinate system is due to the fact that some charges come to this point while others leave it; the charges themselves do not change. But for the time being our solution takes into account only the potential of those charges that are located at the origin.

Very close to the origin, where the retardation becomes a very small quantity, the potential must be determined by the instantaneous value of charge $\delta e(t)$ (see beginning of Section 16).

As was shown in Section 16, the potential of a point charge is equal to $\delta e/r$ (see (6.8)), whence

$$\frac{\Phi(t)}{r} = \frac{\delta e(t)}{r} = \frac{\rho(t)\,dV}{r}$$

Therefore

$$\Phi(t) = \rho(t)\,dV \tag{20.10}$$

The retarded potential of a point charge, $\varphi\,(t - r/c)$, is, in accordance with (20.9) and (20.10),

$$\varphi\left(t - \frac{r}{c}\right) = \frac{\rho(t - r/c)}{r}\,dV \tag{20.11}$$

Now, shifting the coordinate origin to a different point, we obtain an expression similar to (16.9):

$$\varphi\left(t - \frac{|\mathbf{R} - \mathbf{r}|}{c}\right) = \frac{\rho\,(t - |\mathbf{R} - \mathbf{r}|/c,\,\mathbf{r})}{|\mathbf{R} - \mathbf{r}|}\,dV \tag{20.12}$$

Here the charge density is determined at point $\mathbf{r}\,(x,\,y,\,z)$, while the potential is calculated at point $\mathbf{R}\,(X,\,Y,\,Z)$. Thus, the dependence of the charge density on the spatial coordinates is involved in this expression dually: directly, in terms of the argument $\mathbf{r}$, as in the static problem, and in terms of the temporal argument $t - |\mathbf{R} - \mathbf{r}|/c$, owing to the fact that the retardation of the potential at point $\mathbf{R}$ due to different charges of the system is different. Finally, in order to obtain the complete solution of Eq. (20.2), we must integrate (20.12) over all volume elements, that is, over $dV = dx\,dy\,dz$:

$$\varphi = \int \frac{\rho\,(t - |\mathbf{R} - \mathbf{r}|/c,\,\mathbf{r})}{|\mathbf{R} - \mathbf{r}|}\,dV \tag{20.13}$$

For point charges ρ denotes the special function that was defined in Section 12. It is equal to zero at every point except the one in which the charge is located at the given instant.

Equation (20.1) has exactly the same form as (20.2), and its solution satisfies the same boundary conditions. Therefore the vector potential is written analogously to (20.13):

$$\mathbf{A} = \int \frac{\mathbf{j}\,(t - |\mathbf{R} - \mathbf{r}|/c,\,\mathbf{r})}{c\,|\mathbf{R} - \mathbf{r}|}\,dV \tag{20.14}$$

This solution can be compared with (17.10), obtained without taking account of the retardation.

Let us now verify that the solution of the set of equations (20.1) and (20.2) satisfies the Lorentz condition (20.3). The time derivative of the scalar potential is taken directly, because t is involved in the

integral over V as a parameter:

$$\frac{\partial \varphi}{\partial t} = \int \frac{1}{|\mathbf{R}-\mathbf{r}|} \frac{\partial \rho}{\partial t} \, dV$$

The divergence of $\mathbf{A}$ must be taken with respect to the argument $\mathbf{R}$, appearing in the integrand as a parameter, hence it can be taken outside the integral sign; $\mathbf{R}$ is involved only in the combination $|\mathbf{R}-\mathbf{r}|$, so that ∇_R is replaced by $-\nabla_r$. But div_R cannot be directly replaced by $-\mathrm{div}_r$, which refers to the whole expression, because $\mathbf{r}$ is also the spatial argument of $\mathbf{j}$. The following substitution must be made:

$$\mathrm{div}\, \mathbf{R}\, \frac{\mathbf{j}}{|\mathbf{R}-\mathbf{r}|} = -\mathrm{div}_r\, \frac{\mathbf{j}}{|\mathbf{R}-\mathbf{r}|} + \frac{\mathrm{div}_r \mathbf{j}}{|\mathbf{R}-\mathbf{r}|}$$

The divergence of the whole expression (the first term) can be transformed according to the Gauss theorem into a surface integral. By choosing the surface outside the system of currents, we find that the integral vanishes.

In the expression $\mathrm{div}_r\, \mathbf{j}$, the differentiation is only with respect to the spatial argument. Adding $(1/c)(\partial\varphi/\partial t)$ and $\mathrm{div}\, \mathbf{A}$, we get

$$\frac{1}{c}\frac{\partial\varphi}{\partial t} + \mathrm{div}\, \mathbf{A} = \int \frac{1}{c\,|\mathbf{R}-\mathbf{r}|} \left(\frac{\partial\rho}{\partial t} + \mathrm{div}_r\, \mathbf{j} \right) dV$$

But in accordance with the charge conservation law the integrand is zero, so that the Lorentz condition is satisfied.

Retarded Potential at Great Distances From a System of Charges. We shall now look for the form of the solutions of (20.12) and (20.13) at a great distance away from a radiating system. We note that the integrand depends on the argument $\mathbf{R}$ in both integrals in two ways: in the denominator and via the argument $t - |\mathbf{R}-\mathbf{r}|/c$. The function in the denominator depends very smoothly on $\mathbf{R}$. Its expansion in terms of powers of R yields terms which decrease like R^{-n} at infinity. As will be shown later, they add nothing to the radiation for $n > 1$. So we simply replace $|\mathbf{R}-\mathbf{r}|^{-1}$ by R^{-1} and take it outside the integral sign. At large distances from the system the term $|\mathbf{R}-\mathbf{r}|$, appearing in the argument t of the numerator, looks like this:

$$|\mathbf{R}-\mathbf{r}| = R - (\mathbf{r}\cdot\mathrm{grad}\, R) = R - \frac{\mathbf{r}\mathbf{R}}{R} = R - \mathbf{r}\mathbf{n} \tag{20.15}$$

where $\mathbf{n}$ is a unit vector in the direction of $\mathbf{R}$. The subsequent terms of the expansion (20.15) contain R in the denominator and are insignificant. Thus, at a large distance from the radiating system

the potentials are:

$$\varphi = \frac{1}{R} \int \rho \left(t - \frac{R}{c} + \frac{\mathbf{rn}}{c} , \ \mathbf{r} \right) dV \tag{20.16}$$

$$\mathbf{A} = \frac{1}{cR} \int \mathbf{j} \left(t - \frac{R}{c} + \frac{\mathbf{rn}}{c} , \ \mathbf{r} \right) dV \tag{20.17}$$

The term $\mathbf{rn}/c$ in the arguments of the integrands in (20.16) and (20.17) indicates by how much an electromagnetic wave coming from the more distant parts of the radiating system is retarded in comparison with a wave radiated by the nearer parts of the system. In other words, the term $\mathbf{rn}/c$ determines the time that the electromagnetic wave takes to pass through the system of charges. If the velocity of the charges is equal to v, then in that time they are displaced through a distance of $v(\mathbf{r \cdot n})/c$. The retardation inside the system is negligible when this distance is small in comparison with the size of the system r. Therefore, if $v(\mathbf{r \cdot n})/c \ll r$ (or, more simply, $v \ll c$), then the charges do not have time to change their positions noticeably during the time of propagation of the wave in the system.

However, in order that nothing should really change in the system, the charges must also maintain their velocities in that time, because the vector potential depends on the currents, that is, on the particle velocities. If the charges at opposite ends of the system oscillate in phase, and in the time $\mathbf{rn}/c$ the oscillation phase reverses, the action of such charges is mutually weakened owing to retardation within the system.

We shall now formulate the condition in which a retardation of an electromagnetic wave in a system does not result in an additional phase shift between individual emitters. Let the charges oscillate and radiate light of frequency ω. The wavelength of the light is equal to $\lambda = 2\pi c/\omega$. In the time $\mathbf{rn}/c$ the phase of the charge oscillations changes by $\omega(\mathbf{r \cdot n})/c$. This change must be small in comparison with 2π at all points of the system, whence it follows that the size of the system must be small compared with the wavelength of the radiated light in order that the retardation inside the system should not produce a phase shift in waves coming from different emitters.

Thus, the term $\mathbf{rn}/c$ in the argument of the integrand is immaterial, provided two inequalities, $v \ll c$ and $r \ll \lambda$, are fulfilled.

Vector Potential and Field in the Dipole Approximation. Suppose that both inequalities are satisfied, so that the retardation within the system, that is $\mathbf{rn}/c$, is everywhere small. We find that it can be completely neglected only in the expression for the vector potential $\mathbf{A}$; in the scalar expression, using solution (20.16), a higher approximation is necessary. Indeed, after putting $\mathbf{rn}/c = 0$ in (20.16) we find that the integral refers completely to one definite time $t - R/c$,

that is, it no longer depends upon **r** in the time argument. But then we would simply get

$$\varphi = \frac{1}{R} \int \rho \left(t - \frac{R}{c}, \ r \right) dV = \frac{e}{R}$$

where the charge, according to the conservation law, is, in general, time independent. The potential acquires the electrostatic value e/R and makes no contribution to the electromagnetic wave emission effect we are considering.

Instead of making use of subsequent terms in the expansion of the scalar potential in powers of the retardation time within the system, we can eliminate the scalar potential altogether by changing the gauge of the vector potential **A**. Indeed, if the scalar potential does not depend on time, then instead of the Lorentz condition (20.3) we should take

$$\text{div } \mathbf{A} = 0 \qquad (20.18)$$

This does not affect the values of the electromagnetic field components.

Since the gauge of solutions (20.16) and (20.17) corresponded to the Lorentz condition, to meet requirement (20.18) a supplementary term must be added to the vector potential.

It follows directly from Eq. (20.17) that, without taking into account the electromagnetic wave retardation within the system, the vector potential is

$$\mathbf{A} = \frac{1}{Rc} \int \mathbf{j} \left(t - \frac{R}{c}, \ r \right) dV \qquad (20.19)$$

Since here the time argument is not involved in the charge coordinates, we replace **j** by $\rho\mathbf{v}$, remembering that now ρ refers to one time, $t - R/c$. If we are dealing with point charges, the integral of ρ over the domain including the charge is equal to the value of the charge itself. Hence we arrive at the expression

$$\mathbf{A} = \frac{1}{Rc} \left(\sum_i e_i \mathbf{v}^i \right)_{t-R/c} \qquad (20.20)$$

Here $t - R/c$ denotes that the whole sum must be taken at that instant of time. But since $\mathbf{v}^i = d\mathbf{r}^i/dt$, we obtain

$$\mathbf{A} = \frac{1}{Rc} \frac{d}{dt} \left(\sum_i e_i \mathbf{r}^i \right)_{t-R/c} = \frac{\dot{\mathbf{d}}\,(t - R/c)}{Rc} \qquad (20.21a)$$

Here we have used the definition for dipole moment, (16.20). We note that (20.21a) involves only the time derivative $\dot{\mathbf{d}}$. Therefore, the transformation (20.21a), which corresponds to a constant shift

of coordinate origin, does not change $\dot{\mathbf{d}}$ either for a charged system or for a neutral system. In particular, (20.21a) holds also for a single charge.

The approximation (20.21a), in which $\mathbf{A}$ is expressed in terms of a derivative of the dipole moment of the system as a whole, is termed the *dipole approximation*.

In Section 18 a potential gauge transformation for travelling plane waves was chosen such that the scalar potential became zero. We shall make the same gauge transformation for diverging spherical waves, that is, act in accordance with (20.18).

In gauging the potential it should not be differentiated with respect to R, which is involved in the denominator: each such differentiation increases the power of R by unity (because the potential is determined at a great distance from the radiating system). Each time it is sufficient to take the derivative with respect to R, which, being involved in the time argument, defines the retardation. Only terms inversely proportional to R make a contribution to the radiated energy (see further on). The unit vector $\mathbf{r} = \mathbf{R}/R$, which appears in the differentiation, does not have to be differentiated a second time in the required approximation, because in the process superfluous powers of R will appear in the denominator.

We substitute expression (20.21a) into condition (20.18). Applying formula (11.37) for the divergence of a vector depending only on the absolute value of the radius vector, we get

$$\text{div } \frac{\dot{\mathbf{d}} \, (t - R/c)}{Rc} = - \frac{\mathbf{n}\ddot{\mathbf{d}}}{Rc^2}$$

This suggests that the gauge function f should be taken as

$$f = \frac{\mathbf{n}\mathbf{d}}{R} \qquad (20.22)$$

Then, adding to the vector potential the function grad f, we reduce the expression for the vector potential to the form

$$\mathbf{A} = \frac{\dot{\mathbf{d}}}{Rc} - \frac{\mathbf{n} \, (\mathbf{n} \cdot \dot{\mathbf{d}})}{Rc} = - \frac{1}{Rc} \, \mathbf{n} \times (\mathbf{n} \times \dot{\mathbf{d}}) \qquad (20.21b)$$

Here the vector potential is denoted by the same symbol $\mathbf{A}$ as prior to the gauge transformation.

Another operation is also possible: instead of adding to the vector potential grad f, we can find the scalar potential in the next approximation. Then, instead of div $\mathbf{A} = 0$, the equivalent Lorentz condition

$$\text{div } \mathbf{A} + \frac{1}{c} \, \frac{\partial \varphi}{\partial t} = 0$$

is satisfied.

Let us now calculate the electromagnetic field. As pointed out, we need to differentiate only with respect to the argument $t - R/c$. In calculating the magnetic field, we make use of (11.38) and of the fact that curl grad $f = 0$:

$$\mathbf{H} = \text{curl } \mathbf{A} = \frac{1}{Rc} \text{ curl } \dot{\mathbf{d}}\,(t - R/c) = -\frac{1}{Rc^2}\,\mathbf{n} \times \ddot{\mathbf{d}} \qquad (20.23)$$

The electric field is

$$\mathbf{E} = -\frac{1}{c}\frac{\partial \mathbf{A}}{\partial t} = \frac{1}{c}\frac{\partial}{\partial t}\frac{1}{Rc}\,\mathbf{n} \times (\mathbf{n} \times \dot{\mathbf{d}})$$

$$= \frac{1}{Rc^2}\,\mathbf{n} \times (\mathbf{n} \times \ddot{\mathbf{d}}) = \mathbf{H} \times \mathbf{n} \qquad (20.24)$$

From these equations it can be seen that the electric field, the magnetic field, and the vector $\mathbf{n}$ are mutually perpendicular. In
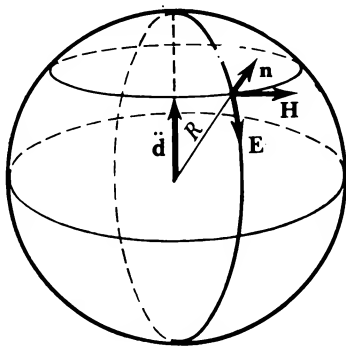
Figure 28

addition, $|\mathbf{H}| = |\mathbf{E}|$, since $|\mathbf{E}|^2 = |\mathbf{H} \times \mathbf{n}|^2 = |\mathbf{H}|^2 - (\mathbf{H}\cdot\mathbf{n})^2$, and $\mathbf{Hn} = 0$. Consequently, the wave at a point R at a great distance from a radiating system is of the nature of a plane electromagnetic wave. This result was to be expected, because the field is calculated far away from charges, where the wave front may be approximately regarded as plane and the solution becomes the same as obtained in Section 18.

Figure 28 gives a general picture of the field. We situate the vector $\ddot{\mathbf{d}}$ at the centre of a sphere of large radius R so that $\ddot{\mathbf{d}}$ coincides with the polar axis or, in other words, is directed towards the "north pole". Through the points where the field vectors are plotted, we draw the "longitude" and "latitude". Then the electric field is tangential to the "longitude" and directed towards the "south", while the magnetic field is tangential to the "latitude" and directed towards the "east". It can be seen from Eqs. (20.23) and (20.24) that the

field vanishes at the "poles" and is maximum at the "equator", that is in a plane perpendicular to $\ddot{\mathbf{d}}$. Thus, the field distribution in space does not possess spherical symmetry. Since the field is transverse and perpendicular to the radius at every point, it cannot be spherically symmetric for purely geometrical reasons. The zone at a large distance from the emitter in which the field is calculated according to Eqs. (20.23) and (20.24) is called the *wave zone*.

The Intensity of Dipole Radiation. Let us now find the energy dissipated by the system in radiation. For this we must calculate the energy flux crossing a sphere infinitely distant from the emitter.

As was shown in Section 15 (see (15.26)), the density of the energy flux, or the Poynting vector, is $(c/4\pi)(\mathbf{E} \times \mathbf{H})$. From this we find that the radiation energy crossing a unit surface at distance R from the radiator in the direction of $\mathbf{n}$ in unit time is

$$\frac{c}{4\pi}\,\mathbf{E}\times\mathbf{H} = \frac{c}{4\pi}\,(\mathbf{H}\times\mathbf{n})\times\mathbf{H} = \frac{c}{4\pi}\,\mathbf{n}\,|\mathbf{H}|^2 \qquad (20.25)$$

The energy flux is directed along the radius from the emitter, as it should be in a wave zone in which the wave is almost plane. The total energy flux crossing the whole sphere of radius R in unit time is

$$\frac{dE}{dt} = \frac{c}{4\pi}\int (\mathbf{E}\times\mathbf{H})\,d\mathbf{S} = \frac{c}{4\pi}\int |\mathbf{H}|^2\,(\mathbf{n}\cdot d\mathbf{S})$$

$$= \frac{c}{4\pi}\int |\mathbf{H}|^2\,dS \qquad (20.26)$$

because $\mathbf{n}$ is directed along $d\mathbf{S}$. Furthermore, the surface element $dS = R^2\,2\pi\sin\vartheta\,d\vartheta$, where ϑ is the polar angle. From (20.23)

$$|\mathbf{H}|^2 = \frac{1}{R^2 c^4}\,|\ddot{\mathbf{d}}|^2\sin^2\vartheta \qquad (20.27)$$

Substituting (20.27) into (20.26), cancelling out R^2, and integrating, we obtain the expression for the energy radiated in unit time:

$$\frac{dE}{dt} = \frac{2}{3}\frac{|\ddot{\mathbf{d}}|^2}{c^3} \qquad (20.28)$$

Note that, for sufficiently large R, none of the terms in the field expressions involving R in the denominator in higher than first power would contribute anything to the energy radiation. That is why only first degree terms in R were retained in the denominator. But if the problem is posed of computing the angular momentum flux rather than the energy flux, the greater accuracy is required in looking for the field at large distances. The angular momentum flux in the wave zone is zero, which can be ascertained from the formulas of Section 15.

Equation (20.28) expresses a result of fundamental importance: whenever a charge is accelerated it radiates energy. Indeed, $\ddot{\mathbf{d}} = \sum e_i \ddot{\mathbf{r}}^i$. Hence, for $\ddot{\mathbf{d}}$ to differ from zero, accelerated motion of the charges is necessary, irrespective of the sign of the accelerations.

Application of this result to the hydrogen atom reveals a striking contradiction with experiment. An electron moving finitely along any orbit must necessarily dissipate all its energy to the electromagnetic field and ultimately fall on the nucleus. Actually nothing of the sort occurs—the atom is stable.

This example reveals how completely inapplicable Newtonian mechanics is to the motion of the electron in an atom. In Part III we shall explain the stability of atoms with the help of quantum mechanics, in which the very concept of motion differs from the classical (Newtonian) view.

Radiation Damping. We have thus found that a charge moving with acceleration continuously radiates energy, transmitting it to the electromagnetic field. We shall now show that, in the case of finite motion of a charge, the dissipation of its energy as radiation can be reduced to the action of a certain effective force of "friction", $\mathbf{F}$. We define this force by the equation

$$\frac{dE}{dt} = -\mathbf{F}\dot{\mathbf{r}} \tag{20.29}$$

since the product of force times velocity is the work done in unit time. Equation (20.28) can be transformed as follows:

$$\frac{dE}{dt} = \frac{2}{3}\frac{e^2}{c^3}|\ddot{\mathbf{r}}|^2 = \frac{2}{3}\frac{e^2}{c^3}\frac{d}{dt}\ddot{\mathbf{r}}\circ\dot{\mathbf{r}} - \frac{2}{3}\frac{e^2}{c^3}\dddot{\mathbf{r}}\circ\dot{\mathbf{r}}$$

We average the obtained relationship over time, taking into account that in finite motion the mean value of a total derivative is zero (see Sec. 17). Then from a comparison with (20.29) we find that the *radiative reaction force* is

$$\mathbf{F} = \frac{2}{3}\frac{e^2}{c^3}\dddot{\mathbf{r}} \tag{20.30}$$

It is proportional to the third time derivative of the radius vector of the charge, or to the second time derivative of its velocity.

By means of somewhat more involved computations it can be shown that the applicability of Eq. (20.30) is not restricted to a finite motion of a charge. The obtained expression is not only the mean but also the instantaneous value of the force of *radiation damping*.

In developing formula (20.28) we neglected the retardation of electromagnetic waves within the system, and the expression is

therefore to a certain degree approximate. But in dealing with a point charge, taking into account the extra power of rn in the integrand in Eq. (20.17) yields a zero contribution at the limit, since for a point charge r tends to zero. For that reason Eqs. (20.28) and (20.30) are strictly applicable to point charges.

Let us find the form of Eqs. (20.28) and (20.30) for the case of fast moving charges. For this we must obtain relativistically invariant formulas, which in the limit of small velocities transform into (20.28) and (20.30). We rewrite the first equation as follows:

$$dE = \frac{2}{3}\frac{e^2}{c^3}\left|\frac{d^2\mathbf{r}}{dt^2}\right|^2 dt$$

Here t is the time in the charge's proper reference frame. In passing to an arbitrary reference frame we must express it, as we know from Section 13, in terms of an invariant quantity, the interval ds:

$$dt = \frac{ds}{c}$$

Further, in four-dimensional notation, $-dE = ic\,dp_4$, $-dt = (i/c)\,dx_4$. Besides, $(d^2r/dt^2)^2 = (du_i/ds) \times (du_i/ds)$, where $u_i = dx_i/ds$, that is, four-dimensional velocity. Hence, in passing to an arbitrary reference frame, in place of the fourth components p_4 and x_4 we must write equations for all the components. This yields

$$dp_i = \frac{2}{3}\frac{e^2}{c}\frac{du_k}{ds}\frac{du_k}{ds}\,dx_i = \frac{2}{3}\frac{e^2}{c}\left(\frac{du_k}{ds}\right)^2 u_i\,ds \qquad (20.31)$$

Now we can find the four-dimensional expression for the force F_i. We start with defining the four-vector of force:

$$m\frac{du_i}{ds} \equiv F_i \qquad (20.32)$$

In the case of a Lorentz force, $F_i = (e/c)\,F_{ik}u_k$. Multiplying both sides of Eq. (20.32) by u_i, we obtain in the left-hand side

$$u_i\frac{du_i}{ds} = \frac{1}{2}\frac{d}{ds}u_i^2$$

But $u_i^2 = -1$, so that $u_i(du_i/ds) = 0$. From this we obtain the condition which any four-dimensional force F_i must satisfy identically:

$$F_i u_i = 0 \qquad (20.33)$$

In addition, we must require that at the limit of low velocities F_i transforms into the three-dimensional nonrelativistic expression (20.31).

As was just shown, the second derivative with respect to time is expressed in terms of the second derivative with respect to the interval. Using the four-dimensional vectors d^2u_i/ds^2 and u_k, we

can develop the one expression which satisfies both requirements:

$$F_i = \frac{2e^2}{3c} \left(\frac{d^2 u_i}{ds^2} + u_i u_k \frac{d^2 u_k}{ds^2} \right) \tag{20.34}$$

Indeed, multiplying it by u_i, we obtain

$$F_i u_i = \frac{2}{3} \frac{e^2}{c} \left(u_i \frac{d^2 u_i}{ds^2} + u_i u_i \times u_k \frac{d^2 u_k}{ds^2} \right) = 0$$

in agreement with (20.33). Furthermore, in the nonrelativistic approximation, $u_\alpha = 0$ and $u_4 = i$, so that $du_4/ds = 0$. Consequently, in such an approximation Eq. (20.34) contains only the first term in the parentheses in the right-hand side, which transforms into (20.31) for three spatial components.

Thus, Eqs. (20.28) and (20.30), together with their relativistic generalizations, should be applicable to arbitrarily moving charges without any restrictions.

But the expression for **F** allows for the accelerated motion of a charge in the absence of any external field! Indeed, assuming **F** to be the only force acting on the charge, we can write the equation

$$m \frac{d^2 \mathbf{r}}{dt^2} = \frac{2}{3} \frac{e^2}{c^3} \frac{d^3 \mathbf{r}}{dt^3}$$

with a solution

$$\mathbf{r} = \mathbf{r}_0 \exp \left(\frac{3mc^3}{2e^2} t \right)$$

corresponding to self-accelerated motion of the charge.

Thus, the electrodynamics of point charges is inherently contradictory.

The difficulties are even greater if a charge is taken to be extended. It becomes necessary to assume the existence of some nonelectric forces which hold the charge together in a finite volume and keep it from flying apart under the action of Coulomb repulsion. But such forces have never been observed. Furthermore, if we assume the existence of a nonelectric force capable of acting on charges, then we must concede that electrodynamics is basically an open-ended theory incapable of resolving questions lying within the domain of its application.

Actually the situation is not all that gloomy. The obtained absurd solution corresponds to self-acceleration of a charge in a time interval of the order of 10^{-23} s. Phenomena which take place in such small time intervals belong to the domain of the quantum theory, where classical notions break down. This will be shown in Part III. For reasons due to quantum theory, classical electrodynamics is applicable to phenomena which take place in not less than 10^{-21} s. This means the following. If a process occurring in, say, 10^{-19} s is examined, the error due to the use of the nonquantum theory is at worst

ot the order of 10^{-2}, while the error associated with the contradiction within the theory itself is of the order of 10^{-4}. In all cases this "intrinsic inaccuracy" is smaller by a factor of one hundred. That is why there is no sense in attempting to change the theory in a way that would eliminate small errors while leaving big ones.

The quantum theory also leads to certain difficulties, but they can be overcome without introducing interactions of a nonelectromagnetic nature.

Magnetic Dipole and Quadrupole Radiation. We have indicated that charges must be in accelerated motion to radiate. But this condition, though necessary, is not sufficient. A simple example can be given when (20.28) yields zero even for accelerated charges. Let a system consist of two identical, charged and colliding particles. According to Newton's Third Law, their accelerations are equal and opposite in sign, so that $\ddot{\mathbf{d}} = \sum e\ddot{\mathbf{r}}_i = e(\ddot{\mathbf{r}}_1 + \ddot{\mathbf{r}}_2) = 0$. In this case the law is applicable because, to the dipole approximation, the retardation of electromagnetic interactions inside the system is considered to be negligible and, hence, the interaction forces between the charges are regarded as instantaneous. But there is then no need to take account of the momentum transmitted to the field, and the total momentum of the particles is conserved, thereby leading to the condition $\ddot{\mathbf{r}}_1 = -\ddot{\mathbf{r}}_2$. For this case Eq. (20.19) yields zero, so that it turns out that the waves emitted by each charge separately mutually cancel out, and it becomes necessary to use higher-order approximations.

If in the expansion of the vector potential (20.17) in powers of **rn** a further term is retained in addition to the zero-power term, we obtain an expression which describes radiation that depends on the change in the quadrupole and magnetic moments of the system. But this expansion is not in inverse powers of R, as in electrostatics and magnetostatics, but in powers of the retardation within the system. All corrections to the radiation field we shall now be looking for are inversely proportional to the first power of R.

Write the corresponding term of the vector potential expansion:

$$\mathbf{A}' = -\frac{1}{Rc^2} \int (\mathbf{r} \cdot \mathbf{n}) \frac{\partial \mathbf{j}}{\partial t} \, dV = -\frac{1}{Rc^2} \frac{\partial}{\partial t} \int \mathbf{j} (\mathbf{r} \cdot \mathbf{n}) \, dV \quad (20.35)$$

The derivative with respect to time is taken outside the integral sign, time being an independent variable. Transform the integral as follows:

$$\int (\mathbf{r} \cdot \mathbf{n}) \mathbf{j} \, dV = \frac{1}{2} \int [(\mathbf{r} \cdot \mathbf{n}) \mathbf{j} - (\mathbf{n} \cdot \mathbf{j}) \mathbf{r}] \, dV$$

$$+ \frac{1}{2} \int [(\mathbf{r} \cdot \mathbf{n}) \mathbf{j} + (\mathbf{n} \cdot \mathbf{j}) \mathbf{r}] \, dV \quad (20.36)$$

Here the first integral is

$$\frac{1}{2} \int [\mathbf{n} \times (\mathbf{r} \times \mathbf{j})]\, dV = \frac{1}{2} \int \rho\, [\mathbf{n} \times (\mathbf{r} \times \mathbf{v})]\, dV = c\mathbf{n} \times \boldsymbol{\mu}$$

that is, it is expressed in terms of the magnetic moment of the system (see (17.19)).

Subtract from the second integral the expression $(2/3)\mathbf{n} \int (\mathbf{r} \cdot \mathbf{j})\, dV$. It corresponds to a vector along $\mathbf{n}$, that is, along the propagation direction of the wave at the given point. According to the general theory, such a component of the vector potential does not affect the field. After this, the second term is equal to

$$\frac{1}{2} \int \rho \left[(\mathbf{r} \cdot \mathbf{n})\, \mathbf{v} + (\mathbf{v} \cdot \mathbf{n})\, \mathbf{r} - \frac{2}{3}\, \mathbf{n}\, (\mathbf{r} \cdot \mathbf{v}) \right]_\alpha dV$$

$$= \frac{1}{2} \sum_i e_i \left[v_\alpha^i\, (\mathbf{n} \cdot \mathbf{r}^i) + r_\alpha^i\, (\mathbf{n} \cdot \mathbf{v}^i) - \frac{2}{3}\, \mathbf{n}_\alpha\, (\mathbf{r}^i \cdot \mathbf{v}^i) \right] \quad (20.37)$$

The obtained result is the time derivative of the quadrupole moment multiplied by vector $\mathbf{n}$ (see (16.18)). If we define the vector

$$Q_\alpha \equiv q_{\alpha\beta} n_\beta \qquad (20.38)$$

then the expression for (20.37) takes the form

$$\frac{1}{2} \int \rho \left[(\mathbf{r} \cdot \mathbf{n})\, \mathbf{v} + (\mathbf{v} \cdot \mathbf{n})\, \mathbf{r} - \frac{2}{3}\, \mathbf{n}\, (\mathbf{r} \cdot \mathbf{v}) \right]_\alpha dV = \frac{d}{dt}\, Q_\alpha \qquad (20.39)$$

Collecting now the obtained expressions, we can find the required term of the vector potential that describes the retardation of an electromagnetic wave within a system of radiating charges:

$$\mathbf{A}' = -\frac{1}{Rc}\, (\mathbf{n} \times \dot{\boldsymbol{\mu}}) - \frac{1}{Rc^2}\, \ddot{\mathbf{Q}} \qquad (20.40)$$

It contains the first derivative of the magnetic moment of the system and the second derivative of the electric quadrupole moment. Strictly speaking, (20.40) should additionally be gauged in accordance with condition (20.18). But this need not be done, bearing in mind that the expression for the magnetic field is not affected by this, while from (20.24) the electric field equals $\mathbf{H} \times \mathbf{n}$, as in any plane or nearly plane wave.

It was already pointed out that when $v/c \ll 1$ and $r/\lambda \ll 1$, the retardation within the system is small. The ratio v/c is involved in the magnetic moment of the system, therefore the terms in the expansion in powers of the retardation that are proportional to v/c yield magnetic dipole radiation. The quadrupole moment of the system contains an additional power of r in comparison with the dipole moment, therefore quadrupole radiation is associated with expansion terms proportional to r/λ.

The field of a magnetic dipole emitter is similar to the field of a radiating electric dipole. Unlike the field repre r nted in Figure 28, a magnetic field radiated by a magnetic dipole lies in the same plane with $\ddot{\boldsymbol{\mu}}$, that is, it is directed along the "longitude", while the electric field is along the "latitude". The formula for the energy radiated in this case is fully similar to (20.28), only instead of $|\ddot{\mathbf{d}}|^2$ it contains $|\ddot{\boldsymbol{\mu}}|^2$. Since the magnetic moment is proportional to v/c, the radiated energy decreases $(v/c)^2$ times in comparison with electric dipole radiation.

The field of a radiating quadrupole is of a more complex shape. The expression for the radiated energy involves the square of the third derivative of the quadrupole moment of the system. In respect of the order of magnitude, a quadrupole radiator emits $(r/\lambda)^2$ less energy in unit time than a dipole emitter.

The electric and magnetic fields of an arbitrary radiating system calculated to the accuracy of the first power of the retardation $(\mathbf{r} \cdot \mathbf{n})/c$ consist of three components: *electrodipole*, *magnetodipole*, and *electroquadrupole*. Accordingly, the Poynting vector contains mixed terms, that is, products of fields corresponding to emitters of different types. But the total radiated energy per unit time is composed of three terms due to each type of radiation separately. This is readily demonstrated as follows.

The radiated energy is a scalar. The mixed terms due to different types of radiation ought to involve scalar combinations of the respective moments that are linear with respect to each of them. Since integration is performed over all directions, that is along vector $\mathbf{n}$, the result can depend only on the product of two moments. But neither of them is a true scalar. Indeed, the product $(\ddot{\mathbf{d}} \cdot \ddot{\boldsymbol{\mu}})$ is obtained from a vector $\ddot{\mathbf{d}}$ and a pseudovector $\ddot{\boldsymbol{\mu}}$ (see Sec. 15), that is, it is a pseudoscalar, which cannot be equal to the radiated energy (a true scalar). Furthermore, $\ddot{d}_\alpha$ and $\ddot{\mu}_\alpha$ have one tensor index each, whereas the quadrupole moment, $\dddot{q}_{\alpha\beta}$, has two tensor indices. Hence a scalar linear in $\dddot{q}_{\alpha\beta}$ and $\ddot{d}_\alpha$ or in $\dddot{q}_{\alpha\beta}$ and $\ddot{\mu}_\alpha$ cannot be developed from $\dddot{q}_{\alpha\beta}$ and $\ddot{d}_\alpha$ or $\ddot{\mu}_\alpha$.

There remains the sum of three terms, each of which is a function of the square of $\ddot{d}_\alpha \ddot{d}_\alpha$, $\ddot{\mu}_\alpha \ddot{\mu}_\alpha$, or $\dddot{q}_{\alpha\beta} \dddot{q}_{\alpha\beta}$. Since the field of a radiating magnetic dipole is similar to the field of a radiating electric dipole, the total energy radiated per unit time by a magnetic dipole is given by a formula similar to (20.28):

$$\frac{dE_q}{dt} = \frac{2}{3c^3} |\ddot{\boldsymbol{\mu}}|^2 \qquad (20.41)$$

It is more difficult to calculate the radiation of an electric quadrupole. Its magnetic field is

$$\mathbf{H}_q = \frac{1}{Rc^3}\, \mathbf{n} \times \overset{\text{\tiny\textbullet\textbullet\textbullet}}{\mathbf{Q}} \tag{20.42}$$

Since in a wave zone the electric field is associated with the magnetic field by the relationship $\mathbf{E} = \mathbf{H} \times \mathbf{n}$, we can write, by analogy with (20.26)

$$\frac{dE_q}{dt} = \frac{1}{4\pi c^5} \int d\Omega\, (\mathbf{n} \times \overset{\text{\tiny\textbullet\textbullet\textbullet}}{\mathbf{Q}})^2$$

$$= \frac{1}{4\pi c^5} \int d\Omega\, [\,|\overset{\text{\tiny\textbullet\textbullet\textbullet}}{\mathbf{Q}}|^2 - (\mathbf{n}\cdot\overset{\text{\tiny\textbullet\textbullet\textbullet}}{\mathbf{Q}})^2] \tag{20.43}$$

From this, using (20.38), we find

$$\frac{dE_q}{dt} = \frac{1}{4\pi c^5} \int \{(n_\alpha \overset{\text{\tiny\textbullet\textbullet\textbullet}}{q_{\alpha\beta}})(n_\nu \overset{\text{\tiny\textbullet\textbullet\textbullet}}{q_{\nu\beta}}) - (n_\alpha n_\beta \overset{\text{\tiny\textbullet\textbullet\textbullet}}{q_\alpha}\, \overset{\text{\tiny\textbullet\textbullet\textbullet}}{q_\beta})^2\}\, d\Omega$$

The integrals over the angles are computed in the following way. The integral of the product of two components n_α and n_ν is other than zero only when $\alpha = \nu$. Hence the integral should be

$$\int n_\alpha n_\nu\, d\Omega = a\delta_{\alpha\nu}$$

To determine a we sum over α. Since $n_\alpha n_\alpha = 1$ and $\delta_{\alpha\alpha} = 3$, we obtain $a = 4\pi/3$.

The integral of the product of four components n_α, n_β, n_ν, and n_γ can be other than zero only when the subscripts of the components are equal in pairs in one of three ways: $\alpha = \beta$, $\nu = \xi$; $\alpha = \nu$, $\beta = \xi$; $\alpha = \xi$, $\beta = \nu$. Hence

$$\int n_\alpha n_\beta n_\nu n_\xi\, d\Omega = b\, (\delta_{\alpha\beta}\delta_{\nu\xi} + \delta_{\alpha\nu}\delta_{\beta\xi} + \delta_{\alpha\xi}\delta_{\beta\nu})$$

To find b we sum again over α. Then we obtain $\int n_\nu n_\xi\, d\Omega$ in the left-hand side and $5\delta_{\nu\xi}b$ in the right-hand side, so that $b = 4\pi/15$. We finally find that

$$\frac{dE_q}{dt} = \frac{1}{5c^5} \overset{\text{\tiny\textbullet\textbullet\textbullet}}{q_{\alpha\beta}}\, \overset{\text{\tiny\textbullet\textbullet\textbullet}}{q_{\alpha\beta}} \tag{20.44}$$

It was stated before that electrodipole radiation cannot occur in collisions of two identical charges. Since angular momentum $\mathbf{M}$ is conserved in collisions, the magnetic moment proportional to it is also conserved, and $\overset{\text{\tiny\textbullet\textbullet}}{\boldsymbol{\mu}} = 0$. Consequently only electric quadrupole radiation remains.

EXERCISES

1. Calculate the time that it takes a charge moving in a circular orbit around a centre of attraction to fall on the centre as a result of the radiation of electromagnetic waves.

2. A particle with charge e and mass m passes, with velocity v, a fixed particle of charge Ze, at a distance ρ. Ignoring the distortion in the orbit of the oncoming particle, calculate the energy that this particle loses in radiation.

Answer.

$$\Delta E = \frac{2}{3} \frac{e^2}{c^3} \int_{-\infty}^{\infty} |\ddot{\mathbf{r}}|^2 \, dt = \frac{2}{3} \frac{Z^2 e^6}{m^2 c^3} \int_{-\infty}^{\infty} \frac{dt}{(\rho^2 + v^2 t^2)^2} = \frac{\pi}{3} \frac{Z^2 e^6}{m^2 c^3 \rho^3 v}$$

3. A plane light wave falls on a free electron, causing it to oscillate. The electron begins to radiate secondary waves, that is, it scatters the radiation. Find the scattering cross section, defined as the ratio of the energy scattered in unit time to the flux density of the incident radiation.

Solution. We proceed from the fact that $\ddot{\mathbf{r}} = e\mathbf{E}/m$. It is apparent from this that if the scattered light propagates perpendicular to the incident beam, scattered radiation is caused only by the component of $\mathbf{E}$ in the third normal direction. The scattered light turns out to be plane polarized. In other directions the light is partially polarized. Knowing $\ddot{\mathbf{r}}$, we determine dE/dt from (20.28). Dividing by the energy flux $c|\mathbf{E}|^2/(4\pi)$, we obtain

$$\sigma = \frac{8\pi}{3} \frac{e^4}{m^2 c^4}$$

4. Find the motion and radiation of a charge elastically connected with a point in space so that the frequency of its free oscillations is ω_0. The charge is located in a magnetic field

$$H_z = |\mathbf{H}|, \qquad H_x = 0, \qquad H_y = 0$$

Solution. The oscillations of the charge are described by the equations

$$m\ddot{x} = -m\omega_0^2 x + \frac{e}{c} |\mathbf{H}| \dot{y}$$

$$m\ddot{y} = -m\omega_0^2 y - \frac{e}{c} |\mathbf{H}| \dot{x}$$

$$m\ddot{z} = -m\omega_0^2 z$$

The third equation is independent of both the first two equations and the magnetic field. The first two equations are easily solved if we put $x = ae^{i\omega t}$

and $y = be^{i\omega t}$. Then

$$a(\omega_0^2 - \omega^2) - i\omega \frac{e|H|}{mc} b = 0$$

$$b(\omega_0^2 - \omega^2) + i\omega \frac{e|H|}{mc} a = 0$$

Now multiply the second equation by i, subtract it once from the first equation and add it once to the first equation. Then the combinations $a \pm ib$ satisfy the equations

$$(a \pm ib)(\omega_0^2 - \omega^2) \mp (a \pm ib)\, \omega \frac{e|H|}{mc} = 0$$

Cancelling out $(a \pm ib)$, we arrive at the frequency equations:

$$\omega^2 - \omega_0^2 \pm \frac{e|H|\,\omega}{mc} = 0$$

Assuming $e|H|/(mc)$ to be small in comparison with ω_0, we replace ω by ω_0 in the term $e|H|\omega/(mc)$ and represent the difference $\omega_0^2 - \omega^2$ as $(\omega_0 + \omega)(\omega_0 - \omega)$, which is approximately equal to $2\omega_0(\omega_0 - \omega)$. Then for the frequencies of both oscillations we obtain

$$\omega = \omega_0 \mp \frac{e|H|}{2mc} \equiv \omega_0 \mp \omega_L$$

They differ from the unshifted frequency ω_0 precisely by $e|H|/(2mc)$, that is, by Larmor's frequency.

After substitution of $\omega = \omega_0 \pm e|H|/(2mc)$ in the equations for a and b, we have, in the same approximation,

$$a = \pm ib$$

Representing the coordinates in real form, we obtain for both oscillations: $x = a \cos(\omega_0 \mp \omega_L)t$, $y = a \sin(\omega_0 \pm \omega_L)t$.

Hence, the radius vector of normal oscillations of frequency $\omega_0 + e|H|/(2mc)$ rotates clockwise, while the radius of the oscillation with frequency $\omega_0 - e|H|/(2mc)$ rotates counterclockwise. Thus, in agreement with Larmor's theorem, the frequency $e|H|/(2mc)$ is either added to the frequency ω_0 or subtracted from it, depending on the sense of rotation of the charge.

Let us investigate the radiation of such a charge in a magnetic field. The electric vector of a radiated electromagnetic wave lies in the same plane as the displacement vector of the charge (see Figure 28). If the radiation is due to the component of the dipole moment, its polarization vector is directed along the z axis, and the field is proportional to $\ddot{z}$. Hence, such radiation is plane-polarized and has a frequency ω_0. A charge oscillating along a magnetic field with the unshifted frequency ω_0 radiates electromagnetic waves polarized in the plane with the magnetic field. A charge oscillating in this manner does not radiate in the direction of the magnetic field at all.

Oscillations with the frequencies $\omega_0 \pm e \mid H \mid /(2mc)$ make for the radiation of circularly polarized waves, the propagation direction of which must coincide with the direction of the magnetic field. The electric field vector in such waves rotates in the same sense as the radius vector of the charge.

If an emitter is placed in the field of an electromagnet, its radiation can be observed perpendicular to the field, while if a hole is drilled in the core of the magnet, the radiation is parallel to the field. In the former case, the observer records oscillations along the z axis, and one projection of each oscillation with right and left polarization. Therefore, both circularly polarized oscillations radiate plane-polarized waves of frequency $\omega_0 \pm$ $\pm e \mid H \mid /(2mc)$ in that direction. A three-fold splitting of the unshifted frequency in the spectrum occurs.

Two frequencies in the spectrum are manifest when observation is along the field; here both oscillations are circularly polarized. When the observation is carried out at an angle to the field, the oscillations are elliptically polarized, and, in addition, there remains the unshifted frequency.

The computations set forth here give the classical theory of the *Zeeman effect*. Actually, in the case of not very strong magnetic fields an entirely different picture of splitting of spectral frequencies is observed; it is correctly described only by the quantum theory.

By observing the Zeeman effect of celestial bodies it is possible to judge the magnitude and direction of the magnetic field near them.

5. Find the radiation energy dissipated in unit time by a charge rotating in an external constant and homogeneous field with a velocity close to that of light.

Solution. We apply formula (20.31). The quantity d^2x_k/ds^2 should be taken from (14.33):

$$\frac{d^2x_k}{ds^2} = \frac{e}{mc} F_{ki} u_i$$

Taking into account that only one component of the magnetic field is other than zero, for example, $H_z = F_{12} = -F_{21}$, we obtain

$$\left(\frac{d^2x_k}{ds^2}\right)^2 = \frac{e^2 \mid H \mid^2}{m^2 c^2}(u_1^2 + u_2^2) = \frac{e^2 \mid H \mid^2 v^2}{m^2 c^2}\left(\frac{dt}{ds}\right)^2 = \frac{e^2 \mid H \mid^2 v^2 E^2}{m^4 c^8}$$

Hence the energy loss in unit time is

$$\frac{dE}{dt} = \frac{2}{3}\frac{e^4 \mid H \mid^2 v^2 E^2}{m^4 c^9}$$

6. Develop the field of a radiating quadrupole in the **wave** zone.

Solution. Let the z axis be directed along the third principal axis of tensor $\overset{...}{q}_{\alpha\beta}$, which we denote simply $D_{\alpha\beta}$. Then the components of vector Q_α, defined by Eq. (20.38), are given by the equations

$$\overset{...}{Q}_x = n_x D_1, \qquad \overset{..}{Q}_y = n_y D_2, \qquad \overset{.}{Q}_z = -n_z(D_1 + D_2)$$

The magnetic field projections in Cartesian components are:

$$H_x = (\mathfrak{n} \times \ddot{Q})_x = -n_y n_z (D_1 + 2D_2)$$

$$H_y = (\mathfrak{n} \times \ddot{Q})_y = n_x n_z (2D_1 + D_2)$$

$$H_z = (\mathfrak{n} \times \ddot{Q}_z) = n_x n_y (D_2 - D_1)$$

From Cartesian components we must pass to spherical components r, ϑ, and φ using an easily developed array of cosines:

	r	ϑ	φ
x	$\sin \vartheta \cos \varphi$	$\cos \vartheta \cos \varphi$	$-\sin \varphi$
y	$\sin \vartheta \sin \varphi$	$\cos \vartheta \sin \varphi$	$\cos \varphi$
z	$\cos \vartheta$	$-\sin \vartheta$	0

and noting that

$$n_x = \sin \vartheta \cos \varphi, \qquad n_y = \sin \vartheta \sin \varphi, \qquad n_z = \cos \vartheta$$

Hence

$$\begin{aligned}
H_r &= H_x n_x + H_y n_y + H_z n_z \\
&= -n_x n_y n_z (D_1 + 2D_2) + n_x n_y n_z (2D_1 + D_2) \\
&\qquad\qquad\qquad\qquad\qquad + n_x n_y n_z (D_2 - D_1) = 0
\end{aligned}$$

as it should be in a wave zone. For the other two components of the magnetic field we have:

$$\begin{aligned}
H_\vartheta &= H_x \cos \vartheta \cos \varphi + H_y \cos \vartheta \sin \varphi - H_z \sin \vartheta \\
&= (D_2 - D_1) \sin \vartheta \sin \varphi \cos \varphi \\
H_\varphi &= -H_x \sin \varphi + H_y \cos \varphi \\
&= (D_1 + D_2 + D_1 \cos^2 \varphi + D_2 \sin^2 \varphi) \sin \vartheta \cos \vartheta
\end{aligned}$$

For $D_2 = D_1$ we obtain for a single-axis quadrupole $H_\vartheta = 0$, as it should be according to symmetry considerations. At $\varphi = 0$ and $\varphi = \pi/2$, that is, on two perpendicular "longitudes", $H_\vartheta = 0$; on the poles and the equator $H_\varphi = 0$.

PART III

QUANTUM MECHANICS

21

THE INADEQUACY OF CLASSICAL MECHANICS. THE ANALOGY BETWEEN CLASSICAL MECHANICS AND GEOMETRICAL OPTICS

The Instability of the Atom According to Classical Notions. Rutherford's experiments established that the atom consists of light negative electrons and a heavy positive nucleus, very small in size in comparison with the atom itself. For such a system to be stable the electrons must of necessity revolve about the nucleus, just as the planets revolve about the sun: opposite charges at rest would be immediately drawn to one another.

But this stability condition is quite insufficient. For one, in a gas the atoms collide continuously, whereas in condensed bodies (liquids or crystals) they are in constant close contact. It is hard to imagine how the atoms of every element retain their identity in such conditions. If, for instance, the solar system were to collide with another stellar system, the state of affairs in both systems would change after the collision.

Besides, as pointed out in the preceding section, the electron's motion in the atom would inevitably lead to radiation of electromagnetic waves, and the dissipation of energy in the process would just as inevitably cause the electron to fall on the nucleus. This, of course, is in striking contradiction with the obvious fact of atomic stability.

Bohr's Theory. In 1913, Niels Bohr suggested a compromise as a way out of this difficulty. According to Bohr, an atom has stable orbits such that an electron moving in them does not radiate electromagnetic waves. But in making a transition from an orbit of higher energy to one with lower energy, an electron radiates, and the fre-

quency of this radiation is related to the difference between the energies of the electron in these two orbits by the equation

$$h\omega = E_1 - E_2$$

where h is a universal constant equal to 1.054×10^{-27} erg-s.

Both of Bohr's principles were in the nature of postulates. But it was possible with their aid to explain, in excellent agreement with experiment, the observed spectrum of the hydrogen atom and also the spectra of a series of atoms and ions similar to the hydrogen atom (for example, the positive helium ion, which consists of a nucleus and one electron). Despite the fact that both of these quantum postulates of Bohr were completely alien to classical physics and could in no way be explained on the basis of classical concepts, they represented an extraordinary step forward in the theory of the atom.

The first postulate states that not every state of the atom is stable, but only certain states. This, as we now know, is in agreement with observations and derives from quantum mechanics just as directly as elliptical planetary orbits derive from Newtonian mechanics.

Bohr's theory was very successful in explaining the spectra of hydrogen and similar atomic systems. But the very next step, a two-electron atom, such as the helium atom, did not yield to consistent calculation by Bohr's theory. The theory was even less capable of explaining the stability of the hydrogen molecule. For this reason, the situation in physics, notwithstanding a number of brilliant results of the Bohr theory, was completely unsatisfactory. Besides the particular difficulties that we have noted here, the Bohr theory was, on the whole, eclectic, since it was inconsistent in its combination of classical and quantum concepts.

Light Quanta. The inadequacy of classical physics for an understanding of physical facts was apparent in very many other cases besides the issue of the stability of the atom. As far back as in 1900, Max Planck demonstrated that the state of thermal equilibrium between field and matter could not be satisfactorily described with the help of the classical laws of radiation, which treat it as a continuous process. In fact, Planck was driven to the assumption that radiators transmit energy to electromagnetic fields in discrete portions, or *quanta*. Each quantum possesses energy in proportion to its frequency, the proportionality factor being the constant h mentioned before. In fact, Bohr made use of Planck's assumption in formulating his second postulate.

The hypothesis concerning light quanta proved extremely useful in explaining a wide range of phenomena. Of great importance was the quantum explanation of the photoelectric effect proposed by Einstein. When the surface of a metal bordering on vacuum is illu-

minated, electrons are ejected from the metal. The energy of each separate electron is found to be quite independent of the total energy of the incident radiation; it depends only on its frequency. The energy of an ejected electron is represented as the difference between two quantities: the energy quantum, $h\omega$, and the work required to remove the electron from the metal.

Expanding on Planck's ideas, Einstein assumed that electromagnetic radiation is not only emitted and absorbed in batches, but also propagates as discrete quanta.

Since the energy of a quantum is equal to $h\omega$, and its speed is c, its momentum should be $h\omega/c$ (see Secs. 14 and 18). Hence, a quantum is a particle of zero mass, the possibility of the existence of which follows from relativistic mechanics.

The momentum of quanta was discovered in the *Compton effect*. In terms of the classical theory, the scattering of electromagnetic waves by free electrons should appear as follows: the incident wave makes an electron oscillate, which causes it to become a radiator in its own right (see Exercise 3, Section 20). But then the frequency of the radiation scattered by the electron must coincide with the frequency of the incident electromagnetic waves.

For an electron within matter to be treated as free the frequency of the incident radiation must be very high in comparison with the electron's natural frequency. Therefore scattering of short-wave radiation (X rays) should be observed. In that case the natural frequency of the electron's oscillations in the optical bend cannot appreciably displace the frequency of the scattered light (that, at least, is what should be expected in accordance with the classical theory).

Reasoning in terms of the principles of classical mechanics, we can assume that the relative shift in the frequency of X rays in scattering on the atomic electrons of matter is the less the higher the oscillation frequency, or the shorter the wavelength. For a given wavelength, scattering at any angle should not result in a change in frequency, since the scattered radiation is emitted by the same oscillating charge in step with the oscillations of the electromagnetic field of the incident wave.

Compton's experiments (1923) revealed an entirely different picture. The harder the incident radiation (that is, the shorter the wavelength) the greater was the observed reduction in the frequency of the scattered radiation (for a given scattering angle). For a specific frequency of the incident radiation, the frequency of the scattered radiation was found to be the lower the greater the scattering angle. A screen through which incident X rays passed freely almost completely blocked rays scattered at a sufficiently large angle.

These results cannot be explained by the classical theory—but they are beautifully explained in terms of quantum notions. The

scattering of X rays on a free electron should be treated as a collision of two particles. One of them—the electron—should be assumed at rest prior to the collision, the other—the light quantum—should be treated as a conventional mass point possessing energy and momentum. The special properties of the quantum are taken into account only by putting its momentum equal to its energy divided by c.

The energy and momentum conservation laws hold in collisions involving quanta in exactly the same way as they do in collisions of any other particles (see Sec. 6). Since in a collision with an electron a quantum imparts a certain momentum to it, its own momentum, and hence its energy, decreases. As a consequence, the frequency of the scattered radiation is lower than the frequency of the incident radiation.

As in any problem on particle scattering, it is sufficient to state one quantity in order to determine all the others. Usually the deflection angle of the initial incident particle is given. We examined a similar problem in Exercise 2, Section 14. Using the result of that problem, and taking into account that the energy of the incident quantum is $E_0 = h\omega_0$, the energy of the scattered quantum is $E = h\omega$, and the initial energy of the electron is mc^2, we obtain

$$\omega = \frac{\omega_0}{1 + h\omega_0 (1 - \cos \theta)/(mc^2)}$$

This gives the dependence of the frequency of the scattered radiation on the scattering angle θ. It agrees fully with experimental data.

The notion of light as a stream of corpuscles was not new in physics, but by the time the hypothesis of light quanta was enunciated the wave theory of light had been generally recognized, while the corpuscular theory seemed abandoned forever. Such phenomena as diffraction and interference of light can, in the classical theory, be interpreted only on the basis of the wave picture and utterly contradict the corpuscular one. On the other hand, the photoelectric and Compton effects are as hopelessly at variance with the classical wave theory.

Thus, by the beginning of the 1920s the science of physics found itself in the unusual position of having to rely on two apparently fundamentally contradictory theories to explain various phenomena involving the same essence, the electromagnetic field. The way out of the situation was provided by consistent quantum theory, in which every type of motion possesses certain wave properties.

Classical Newtonian mechanics, it was found, is restricted to macroscopic bodies, often breaks down when applied to microscopic entities, and is totally unsuitable in describing the motions of

atomic electrons or light quanta. Quantum mechanics embodies the exact criterion of applicability of classical notions.

But before taking up quantum mechanics we must explain how one and the same entity can, as a matter of principle, display both corpuscular and wave properties.

The Correspondence Between Geometrical Optics and Classical Mechanics. The corpuscular theory of light may have been due to the fact that in a homogeneous medium light rays propagate in a a straight line, like particles not subject to the action of the medium. But plane electromagnetic waves propagate in an infinite homogeneous medium just as rectilinearly: the normal to the wave front is constant in direction. That being the case, there exists a a reciprocal correspondence between the corpuscular and wave pictures. The difference between them appears when waves propagate in a restricted space. The one-to-one correspondence between a wave front and the direction of the wave normal is violated by the relationships (19.10). Whereas in the case of a perfectly plane wave at every point in space one definite straight line (the wave normal) can be constructed and it can be compared with the path of a corpuscle, in the case of a "smeared" wave front there is a cone of directions at every given point and no physical means of stating the corpuscle's "actual" path. All directions within the cone are equally valid, but that means that, strictly speaking, none of them are valid.

Constructions with light beams are used in geometrical optics. But such constructions, we have seen, are the more ambiguous the less the precision with which the direction of the wave normal can be defined. And the latter depends on the dimensions of the domain in which the electromagnetic wave propagates. As long as the domain is in all dimensions large in comparison with the wavelength of the light, no diffraction (wave) effects come into play. But as the dimensions of the domain approach the wavelength of light the concept of a light ray becomes increasingly meaningless.

Thus, on the basis of purely wave notions it is possible to establish a criterion for the applicability of the ray concept (that is, the corpuscular picture) and establish the rules for the limiting transition from wave to mechanical concepts. This, in turn, indicates the way to pass from the corpuscular concepts of Newtonian mechanics back, so to say, to the wave concepts of quantum mechanics.

But first it is necessary to establish the correspondence between mechanical and wave quantities.

Surfaces of Constant Phase. First, let us determine the significance of the wave phase in light-ray optics. The very notion of phase would appear to be entirely alien to geometrical optics, but in fact,

as will now be shown, it is possible to give a definite mechanical meaning to the concept of phase.

We shall begin with the expression for the field of a propagating electromagnetic wave in the form

$$\mathbf{E} = \mathbf{E}_0(\mathbf{r}, t) \cos \frac{\chi(\mathbf{r}, t)}{\lambda} \tag{21.1}$$

Here, λ is the wavelength, regarded as small in comparison with the linear dimensions of the domain occupied by the field. In the limiting case of a plane wave the phase is

$$\frac{\chi}{\lambda} = \mathbf{k}\mathbf{r} - \omega t \tag{21.2}$$

(cf. (18.25) and (18.26)). But since $\mathbf{k} = 2\pi\mathbf{n}/\lambda$ and $\omega = 2\pi u/\lambda$, where u is the phase velocity (see (19.7)), it is convenient to exhibit the λ dependence explicitly by writing the phase as the quotient $\varphi \equiv \chi/\lambda$.

The expression for the field in terms of phase must be substituted into the wave equation in order to specify the order of magnitude of all the terms that should be discarded for the transformation to the approximation of ray, or geometrical, optics:

$$\nabla^2 \mathbf{E} - \frac{1}{u^2} \frac{\partial^2 \mathbf{E}}{\partial t^2} = 0 \tag{21.3}$$

In differentiating with respect to t and $\mathbf{r}$, we retain only those terms that contain the highest degree of λ in the denominator, because λ is by definition a small quantity. That is why it is unnecessary to differentiate the amplitude of the wave packet, $\mathbf{E}_0(\mathbf{r}, t)$. Consequently

$$\frac{\partial \mathbf{E}}{\partial t} \approx -\frac{\mathbf{E}_0}{\lambda} \frac{\partial \chi}{\partial t} \sin \frac{\chi}{\lambda}$$

$$\frac{\partial^2 \mathbf{E}}{\partial t^2} \approx -\frac{\mathbf{E}_0}{\lambda} \frac{\partial^2 \chi}{\partial t^2} \sin \frac{\chi}{\lambda} - \frac{\mathbf{E}_0}{\lambda^2} \left(\frac{\partial \chi}{\partial t} \right)^2 \cos \frac{\chi}{\lambda}$$

The first term in the second line should be discarded in comparison with the second term, containing λ^2 in the denominator. Hence

$$\frac{\partial^2 \mathbf{E}}{\partial t^2} \approx -\mathbf{E}_0 \left(\frac{1}{\lambda} \frac{\partial \chi}{\partial t} \right)^2 \cos \frac{\chi}{\lambda}$$

and similarly

$$\nabla^2 \mathbf{E} \approx -\mathbf{E}_0 \left(\frac{1}{\lambda} \operatorname{grad} \chi \right)^2 \cos \frac{\chi}{\lambda}$$

Substituting these expressions into (21.3), we obtain a first-order differential equation for the phase $\varphi = \chi/\lambda$:

$$(\operatorname{grad} \varphi)^2 - \frac{1}{u^2} \left(\frac{\partial \varphi}{\partial t} \right)^2 = 0 \tag{21.4}$$

In the limiting case of a plane wave, it follows from (21.2) that

$$\mathbf{k} = \operatorname{grad} \varphi \equiv \frac{\partial \varphi}{\partial \mathbf{r}} \tag{21.5}$$

and

$$\omega = -\frac{\partial \varphi}{\partial t} \tag{21.6}$$

where for a plane wave

$$k^2 - \frac{\omega^2}{c^2} = 0$$

But according to (21.4) this equation is also satisfied by the quantities grad φ and $-\partial\varphi/\partial t$ in an almost plane wave (21.1). Consequently, Eqs. (21.5) and (21.6) can be taken as definitions of the wave vector and frequency of an almost plane and almost monochromatic wave. (We assumed the duration of the wave process represented by Eq. (21.1) to be much greater than the oscillation period $2\pi/\omega$.)

It is apparent from (21.5) that the wave vector is directed along the normal to the surface of constant phase $\varphi = \varphi_0$, that is, it defines the direction of the light ray at the given point in space. The propagation of an almost plane wave is represented as a spatial displacement of a family of surfaces of constant phase.

At different instants of time t a surface of some specific value $\varphi = \varphi_0$ occupies different positions in space according to the equation

$$\varphi(\mathbf{r},\ t) = \varphi_0$$

Let us determine the speed with which this surface propagates. For that we proceed from the condition

$$d\varphi = \frac{\partial \varphi}{\partial t}\, dt + \frac{\partial \varphi}{\partial \mathbf{r}}\, d\mathbf{r} = 0$$

Let $d\mathbf{r}$ be a vector directed normal to the surface. Then $|\partial\varphi/\partial\mathbf{r}|$ is the absolute value of $\mathbf{k}$. In accordance with the definition of phase velocity (19.7), we obtain from (21.5) and (21.6)

$$\left| \frac{d\mathbf{r}}{dt} \right| = \frac{|\partial\varphi/\partial t|}{|\partial\varphi/\partial\mathbf{r}|} = \frac{\omega}{k} = u$$

The group velocity of propagation of the wave packet (21.1) can be determined with the help of (19.5) as

$$\mathbf{v} = \frac{\partial \omega}{\partial \mathbf{k}} \tag{21.7}$$

It is essential that for an almost plane wave ω can be expressed as a function of $\mathbf{k}$, as is done for a plane wave.

Similar Quantities in the Optico-Mechanical Analogy. In¦ Section 10 we examined the propagation of surfaces of constant action of a system of identical mass points travelling along a bundle of paths. At the initial time the initial conditions for each such particle are stated. As the particles move the value of the action varies for each of them according to the equation

$$S = \int_{t_0}^{t} L \, dt$$

It was established before that the propagation of surfaces $S =$ constant is described by a first-order partial differential equation in the form (10.20), in which we must substitute $V = S$ if the action of the particles itself is taken as the function responsible for the canonical transformation.

The Hamilton-Jacobi equation is fully analogous to the equation of the propagation of a constant-phase surface (21.4). The latter can, for example, be written as follows:

$$[u^2 (\text{grad } \varphi)^2]^{1/2} = -\frac{\partial \varphi}{\partial t}$$

Then the left-hand side of the equation is similar to the Hamiltonian, in which the wave vector $\mathbf{k} = \text{grad } \varphi$ has been substituted for the momentum and the coordinate dependence is involved by means of the quantity u. Equation (21.4) is even more like the Hamilton-Jacobi equation in relativistic form if we substitute $m = 0$ into the Hamiltonian of a free particle (14.12) and replace $\mathcal{H}$ by $-\partial S/\partial t$, and p by grad S.

Thus, the propagation of light rays in a medium is fully analogous to the motion of particles of zero mass. The mechanics of these particles is defined by Eq. (21.4) to exactly the same degree as the mechanics of "conventional" mass points is by the Hamilton-Jacobi equation.

To every quantity in mechanics there corresponds a similar quantity in geometrical optics. The similarity of the quantities can be established on the basis of a comparison between phase and action. It will be observed that frequency corresponds to energy, and the wave vector to momentum. Indeed, the correspondence between E and ω can be seen from (10.26) and (21.6):

$$E = -\frac{\partial S}{\partial t}, \qquad \omega = -\frac{\partial \varphi}{\partial t}$$

while from (10.24) and (21.15) we see the correspondence between $\mathbf{k}$ and $\mathbf{p}$:

$$\mathbf{p} = \frac{\partial S}{\partial \mathbf{r}}, \qquad \mathbf{k} = \frac{\partial \varphi}{\partial \mathbf{r}}$$

But then the propagation velocity of a constant-action surface, according to (10.27), and that of a surface of constant phase have quite similar expressions:

$$\frac{E}{|\mathbf{p}|} \qquad \text{and} \qquad \frac{\omega}{|\mathbf{k}|}$$

Finally, the velocity of a wave packet is analogous to the velocity of the particles:

$$\mathbf{v}_{\text{particle}} = \frac{dE}{d\mathbf{p}}, \qquad \mathbf{v}_{\text{packet}} = \frac{d\omega}{d\mathbf{k}}$$

In Section 18 we found the law for transforming the frequency and wave vector, which coincides with the law for transforming the energy and momentum of a particle in passing from one reference frame to another.

Corresponding optical and mechanical quantities differ only in dimension. Thus, phase has zero dimension, while the dimension of action is that of $\int L\, dt$, that is g-cm^{2}s^{-1}. The wave vector and momentum, frequency and energy also differ correspondingly by the dimension of action. The coefficient of proportionality must be the same in all relationships, because otherwise the optico-mechanical analogy would not be relativistically invariant. This coefficient, as we shall see later, is again the action quantum, or Planck's constant h.

In the following section it will be shown that the optico-mechanical analogy is a consequence of a limiting transition from the wave equations of quantum mechanics to the equations of classical mechanics, similar to the limiting transition from the wave equation of electrodynamics to the equation of propagation of light rays.

EXERCISE

Proceeding from the fact that phase is analogous to action, show that light of given frequency is propagated along paths for which the propagation time of constant phase is least (the Fermat principle).

Solution. At constant frequency

$$\varphi = \int_1^2 \mathbf{k}\, d\mathbf{r} = \omega \int_1^2 u^{-1}\, (\mathbf{n} \cdot d\mathbf{r})$$

The product $(\mathbf{n} \cdot d\mathbf{r})$ is the displacement of the surface in a direction normal to it. It follows that $u^{-1}(\mathbf{n} \cdot d\mathbf{r})$ is the propagation time dt. In accordance with the variational principle, which governs phase as well as the analogous quantity of action, the time $t = \int_1^2 dt$ is the least possible time.

In mechanics a similar principle holds when the energy of the particles is constant. Then the action should be written in the form

$$S = \int \mathbf{p} \, d\mathbf{r}$$

But the momentum $\mathbf{p}$ is usually associated with the velocity of the particle itself ($\mathbf{p} = m\mathbf{v}$) and not with the constant-action surface (the Euler-Maupertuis principle of least action).

22

ELECTRON DIFFRACTION

The Essence of Diffraction Phenomena. Classical mechanics is analogous only to geometrical optics, and by no means to wave optics. The difference between mechanics and wave optics is best illustrated by the example of diffraction phenomena.

Let us consider the following experiment. Let there be a screen with two small apertures. Let us assume that the distance between the apertures is of the same order of magnitude as the apertures themselves. We cover one of the apertures (which we call the "first") and direct a light wave on the screen. We observe the wave passing through the second aperture by the intensity distribution on a second screen situated behind the first. Then we cover the second aperture. The intensity distribution changes. Now we open both apertures together. An intensity distribution is obtained which in no way represents the sum of the intensities due to each aperture separately. At the points of the screen at which the waves from both apertures arrive in opposite phase they mutually cancel, while at those points at which the phase for both apertures is the same, they reinforce each other. In other words, it is not the intensities of light, that is the quadratic values, that are added, but the values of the fields themselves.

This type of diffraction can occur only because the wave passes through both apertures. Only then are definite phase differences obtained at points of the second screen for rays passing through each aperture. (We disregard here the diffraction effects associated with the passage through one aperture. These phenomena are due to the phase differences of the rays passing through various points of the aperture. Instead of examining such phase differences we assume that the phase of the wave passing through each aperture is constant, but we take into account the phase differences between waves passing

through different apertures. Nothing is essentially changed by this simplification.)

A somewhat more complicated picture is that of the diffraction of X rays in a crystal lattice, because the diffraction phenomena take place in three dimensions. It is best to imagine the lattice as a stack of planes filled with individual atoms or molecules. A wave passing through the lattice is partially reflected from each such plane at the same angle. But since the reflections occur at constant distances from each other, a constant phase difference appears between the reflected waves, which depends on the angle of incidence of the wave on the crystal. The reflected waves may be mutually reinforced only if the phase difference is, depending on the angle of incidence, equal to 2π, 4π, 6π and so on.

The distance between the planes in crystals of simple structure can be calculated without resorting to X-ray diffraction (from the density, atomic weight, and the Avogadro number), knowing the number of atoms per cubic centimetre. The angle at which reflection from a given crystallographic plane is possible is related to the distance between the planes by a simple geometric condition involving the wavelength (the Bragg law) so that the wavelength of X rays can be determined by measuring the reflection angle.

Although the three-dimensional diffraction that occurs in a crystal lattice is more complex than the plane picture of diffraction at apertures, the cause of these phenomena is the same: just as a wave must pass through both apertures simultaneously for a phase difference to appear on the screen, the scattering of X rays must occur on all atoms of the crystal lattice. The same wave must be scattered on every atom of the crystal. Then the reflection condition is satisfied only at strictly definite values of the angles. No corpuscular picture could explain X-ray diffraction without taking into account the wave properties of radiation.

Electron Diffraction. The picture is exactly the same in the scattering of electrons (as well as of other microparticles) on crystals. Electrons, as we know, act on a photographic plate or luminescent screen in a way similar to X rays, and direct experiments accordingly reveal that microparticles undergo diffraction governed by the same basic laws as electromagnetic wave diffraction.

But for that each electron must be scattered on all the atoms of the lattice, since electrons travel entirely independently of one another: there can be no constant phase difference between them. They may simply pass through a crystal one by one, and the diffraction picture will be the same as in the passage of all at once.

In optics, it will be recalled, a diffraction pattern can be obtained when only one light source is used—beams from different sources are not coherent. It is necessary for the same light wave to pass

through both apertures: only then can the path difference maintain its constant value, which depends on the geometry of the instruments in the experiment as well as on the wavelength. In this case light and dark spots persist at the same points of the screen despite the fact that the rays emitted by the source at different times are not coherent. But this is of no consequence in observing diffraction: whenever the wave was emitted, on passing through both apertures it reconverges, reinforcing or damping itself on the screen, depending on the number of half-waves that fits into each path from the apertures to the respective points on the screen.

Electron diffraction proves that in the microworld the laws of motion are in general of a wave nature, and that each electron is scattered on all the atoms of a crystal lattice. This is obviously incompatible with the concept of a definite path of an electron, just as X-ray diffraction is incompatible with ray, or geometrical, optics.

Diffraction phenomena are proof that electron motion is associated with a phase of certain magnitude.

Electron Wavelength. Just as the wavelength of X rays is determined according to their diffraction, so the diffraction of electrons (or other microparticles) makes it possible to measure the wavelength associated with them. There exists a very simple relationship between the wavelength and the velocity of the momentum $\mathbf{p}$ of a particle. The wave vector of a particle is associated with its momentum by the relationship

$$\mathbf{k} = \mathbf{p}/h \qquad (22.1)$$

This relationship was enunciated by Louis de Broglie several years before it was experimentally confirmed by C. Davisson and L. Germer, who first observed the diffraction of electrons on crystals.

The constant h, or action quantum, was mentioned in the preceding section. Formerly, a constant 2π times greater than h was commonly used, and the value of h used in this book was denoted $\hbar$. Also, the number of oscillations per second, ν, was used in the expression for phase instead of the number of radians per second, ω.

The wavelength of an electron corresponding to (22.1) is

$$\lambda = \frac{2\pi}{k} = \frac{2\pi h}{p} = \frac{2\pi h}{mv} \qquad (22.2a)$$

and it is called the *de Broglie wavelength*. A quantity **smaller by** a factor of 2π,

$$\lambdabar = \frac{1}{k} = \frac{h}{p} \qquad (22.2b)$$

is often used.

Equation (22.2a) shows that the de Broglie wavelength corresponding to the motion of macroscopic bodies is extremely small, owing to the smallness of h with respect to all the quantities that could characterize the motion of such bodies. If the de Broglie wavelength is expressed in terms of the natural units of macroscopic motion, centimetre-gram-second, for a body of mass, say, one gram and velocity 10^5 cm-s^{-1}, it yields $\lambda \approx 10^{-22}$ cm. Obviously, it is impossible to observe the motion of a macroscopic body in a domain of such small dimensions, hence no diffraction phenomena can be observed. The applicability of classical (Newtonian or relativistic) mechanics to macroscopic bodies is in practice unrestricted.

The Limits of Applicability of Classical Concepts. The relationships between quantities are entirely different when equation (22.2a) is applied to the motion of an electron in an atom. As mentioned before, the dimensions of atoms are easily determined by dividing the volume of one gram atom of a solid or liquid substance by the Avogadro number $N_A = 6.024 \times 10^{23}$. The atomic radius is of the order of 0.5×10^{-8} cm.

The order of magnitude of the velocity of an electron is evaluated as follows. Write the equation of motion of a charge in a Coulomb field

$$m\ddot{\mathbf{r}} = -\frac{Ze^2\mathbf{r}}{r^3}$$

Multiply both sides scalarly by $\mathbf{r}$ and transform the second derivative with respect to time in the left-hand side thus:

$$m\ddot{\mathbf{r}}\mathbf{r} = m\frac{d}{dt}(\dot{\mathbf{r}}\cdot\mathbf{r}) - m\dot{\mathbf{r}}^2 = -\frac{Ze^2r^2}{r^3} = -\frac{Ze^2}{r} = U$$

We obtain the expression for potential energy. Now average the equation over a certain sufficiently large time interval. Then the mean value of the total derivative is in any case zero: as we know from experience in normal conditions the atom is stable. But it follows from this that the mean value of its kinetic energy is equal to one-half the mean potential energy taken with the opposite sign. Substituting the obtained evaluation of the radius, $r = 0.5 \times 10^{-8}$ cm, $Z = 1$ (a hydrogen atom), and $m = 9 \times 10^{-28}$ g (the electron mass), we obtain

$$v = \left(\frac{e^2}{mr}\right)^{1/2} \approx 2.2 \times 10^8 \text{ cm-s}^{-1}$$

Thus, the wave number corresponding to the motion of an electron as defined by Eq. (22.2a) is equal to 0.5×10^{-8} cm, that is, it is a quantity of the dimensions taken for the atom itself.

If we assume that there is some electron path in an atom, then its total length is that of only one de Broglie wavelength. In that

case it is obviously meaningless to speak of a path at all: it must be totally smeared by diffraction phenomena. The picture is as if we were considering a wave packet one wavelength in size. The amplitude of the wave would be other than zero at all points of the packet, so that no curve drawn within the domain of the packet could be correlated with an imagined path of a particle or a light beam. In considering the motion of a particle within such a packet it can be described only in the framework of wave concepts. (This, of course, does not refer to the motion of the packet through space as a whole!)

We thus see that the motion of an electron in an atom is essentially a wave phenomenon. The concept of path loses meaning.

At the same time, it should be remembered that quantum description does not deny the electron its properties as a particle, and it does not become a wave in the conventional sense of the word. For example, a part of an electron is never observed. If the second screen on which the diffraction picture appears is replaced by a photographic plate it will display but one blackened point for every impinging electron. Only the configuration of the blackened points as a whole characterizes the diffraction pattern. It is thus more correct to say not that the electron becomes a wave but that in the microworld the laws of motion are of wave nature.

At the same time, diffraction would have been quite impossible if all atoms of a crystal were not actually involved in the passage of the same electron. An electron path of the form we are used to with conventional macroparticles simply does not exist in a diffraction experiment. What specifically is of a wave nature in this case will be shown later. Whatever the case may be, in no experiment displaying the wave properties of motion is any splitting of electron charge or mass ever observed.

This does not mean that the electron does, after all, possess some definite path which we are simply incapable of observing, owing to the inadequacy of our experimental equipment or physical knowledge. Diffraction phenomena are specific proof that the electron has no path, just as diffracting light does not propagate in separate beams. In the diffraction experiment light passes through both apertures, which is incompatible with the concept of a single beam. In the same sense there is no such thing as an electron's path in an atom: this is a firmly ascertained fact which cannot be refuted by any subsequent development of physics.

Statistical Regularity and an Isolated Experiment. The absence of a path certainly does not mean that there are no regularities in electron motion. Quite the contrary—identical diffraction experiments, involving, of course, a sufficiently large number of electrons having the same velocities, always yield identical diffraction pat-

terns. There is thus no doubt that causal regularity exists, but it is of a statistical nature, appearing in a very large number of separate experiments, since each passage of an electron through a crystal can be regarded as a separate, independent experiment.

Diffraction phenomena produce a regular pattern of dots on a photographic plate in the same way as a large number of shots at a target is subject to a scattering law. However, unlike bullets, which travel along paths and therefore produce a smooth distribution curve of target hits, electrons produce a more erratic pattern of blackened spots, which characterizes wave motion. The scattering of bullet hits is due to the impossibility of exactly reproducing the initial firing conditions and can be reduced by better aiming; electron scattering, on the other hand, produces a regular diffraction pattern which can in no way be changed for a given velocity of the electrons.

It should also be noted that statistical regularities in diffraction experiments have nothing in common with the statistical regularities that govern the motions of large assemblies of interacting particles. As repeatedly stressed, the same picture is observed quite irrespective of how the electrons pass through a crystal: all at once or one by one. A certain phase governing the motion exists only because each electron interferes with itself. Of course, quantum laws of motion also influence the behaviour of large assemblies of particles, affecting the statistical laws inherent in assemblies, but unlike the classical laws of motion, they do not lose their probability nature in going over to individual electrons.

The Uncertainty Principle. It remains now to examine in greater detail the question: In what cases does the concept of path of a microparticle remain meaningful? The paths of particles in the Wilson cloud chamber, the cathode-ray oscillograph and many other instruments can be excellently precalculated according to the laws of classical mechanics, and visually observed. A moving electron in a Wilson cloud chamber actually leaves a very real cloud track.

Let us first recall that under certain conditions light propagates along definite paths, rays or beams. Geometrical optics holds when the inaccuracy in stating the wave vector Δk_x, which is subject to the inequality

$$\Delta k_x \, \Delta x \gtrsim 2\pi \tag{22.3}$$

is small in comparison with k_x (see (19.9)). Substituting Δk_x from Eq. (22.1), we obtain a similar inequality for a microparticle, an electron, for example:

$$\Delta p_x \, \Delta x \gtrsim 2\pi h \tag{22.4a}$$

This inequality is known as the *uncertainty relation* of quantum mechanics.

If we take the evaluation (19.6b) instead of (22.4a), for $\langle\Delta k_x\rangle\langle\Delta x\rangle$, a similar inequality is obtained:

$$\langle\Delta p_x\rangle\langle\Delta x\rangle \gtrsim h \qquad (22.4b)$$

The concept of an electron path has meaning if the uncertainty of all three of its momentum components is small in comparison with the momentum itself:

$$\Delta p_x \ll p_x, \qquad \Delta p_y \ll p_y, \qquad \Delta p_z \ll p_z \qquad (22.5)$$

Note that although we have been using "electron" all along, simply to be specific, any microparticle is implied.

It is easy to see why the uncertainty relations in no way prevent electrons from having paths, for example, in a TV-kinescope. For the sake of the evaluation, let us take the dimensions of an image element to be around 0.1 cm. Then from (22.4b) the uncertainty in the transverse momentum component is 10^{-26} g-cm-s^{-1}. Since the electron velocity here is around 10^{10} cm-s^{-1}, its momentum is approximately 10^{-17}. Hence, the uncertainty in the angle specifying the direction of the momentum is of the order of 10^{-9}, and the inaccuracy with which the electron beam reaches the screen does not exceed 10^{-7} cm, yielding a "safety margin" of $10^{-1}/10^{-7}$, or a million-fold.

This agrees with the conclusion derived for an atom, where no such "safety margin" exists. An atom is 10^7 times smaller, so that $\Delta p \sim 10^{-19}$; the momentum of an electron in an atom, at a speed of 2×10^8, is also of the order of 10^{-19}; that is, Δp and p are of the same order of magnitude, and there can be no electron path in an atom.

Thus, quantum mechanics does not abolish classical mechanics, but includes it as a limiting case, just as wave optics includes the limiting case of the geometrical optics of light rays. Moreover, quantum mechanics deals with the same quantities as classical mechanics: energy, momentum, coordinates, angular momentum; but the finite nature of the action quantum h imposes restrictions on the simultaneous applicability of any two classical concepts (for instance, coordinates and momentum) to one and the same state of motion.

The momentum and coordinates of a particle cannot have precise values simultaneously because of the essentially wave nature of its motion. It is as meaningless to attempt to define their exact values as it is to seek an exact path of light beams in wave optics. Just as light rays cannot be made more precise by improving the optical instruments, no progress in measuring techniques will ever make it possible to determine an electron's path to a greater accuracy than indicated by the relationships (22.4a) or (22.4b). As we have

seen, the very concept of path or trajectory has the same approximate meaning as the concept of a light ray or beam.

Sometimes erroneous attempts are made to interpret uncertainty relationships. For instance, it is assumed that a path cannot be determined because the accuracy of the initial conditions does not exceed Δp_x and Δx, connected by relation (22.4a). This is to say that some actual path does exist somewhere, but it lies in a more or less narrow spatial domain and limited range of momenta. The "real" path is likened to the imaginary trajectory drawn from a gun to a target before firing. The path of the projectile is not precisely known beforehand, if only because strictly identical powder charges cannot be prepared. But this inaccuracy in the initial conditions for a bullet only leads to a smooth scattering curve for the hits on the target, while the distribution of electrons is subject to laws of wave diffraction: there exist minimum and maximum regions in no way associated with inaccurate knowledge of the initial conditions. Diffraction shows that no "real-though-unknown-to-us" trajectory exists.

Uncertainty relations state not the error to which certain quantities can be measured simultaneously, but to what extent these quantities have precise meaning in the given motion. It is this that the uncertainty principle of quantum mechanics expresses. The term "uncertainty" emphasizes the fact that what we are concerned with is not accidental errors of measurement or the imperfection of physical apparatus, but the fact of momentum and coordinate of a particle being actually meaningless for the same state of a microparticle.

23

THE WAVE EQUATION

The Wave Equation. Diffraction occurs because of the superposition of wave amplitudes. When the phases coincide the intensity, which is proportional to the square of the resultant amplitude, is maximum; when the phases are opposite the intensity is minimum. In electron diffraction the quantity analogous to intensity is measured according to the blackening of a photographic plate, that is, the number of electrons impinging on unit area. The alternation of maxima and minima and their relative configurations in electron diffraction are subject to the same law as X-ray diffraction. In order to explain the diffraction of electrons it must be assumed that their motion,

like the propagation of waves, can be associated with some wave function the behaviour of which determines the diffraction pattern.

The blackening of the plate, that is, the number of electrons striking a unit area, is a physically observable quantity. Just as the number of hits in an artillery barrage is proportional to the hit probability in a given grid square, the density of electrons striking the photographic plate is in direct proportion to the probability of their occurring in the vicinity of a given point of the plate. Each electron impact should be regarded as an identical experiment the result of which is not predetermined and must be predicted on a probability basis, by stating the ratio between the probability density and the wave function of the electron's motion.

The analogy with the diffraction of an electromagnetic wave is extremely useful. The blackening of the plate caused by a wave is proportional to the square of the wave amplitude; it can therefore be expected that the probability of an electron occurring at a certain point is proportional to the square of the wave function. Since the observable quantity is the probability density, not the wave function, the latter must in the most general case be treated as a complex quantity the square of the modulus of which must be taken in order to pass to the probability density. Further on we shall see that a real wave function may correspond to far from all states.

We shall assume the probability of an electron (or particle in general) occurring in a volume element dV to be related to the wave function by the following relationship:

$$dw = |\psi(x, y, z, t)| dV \qquad (23.1)$$

where the wave function ψ depends on position and time; the square of its modulus is the probability density.

The Linearity of the Wave Equation. Just as in electrodynamics Maxwell's equations are used to investigate the laws of propagation of a wave itself, and the intensity is found by squaring its amplitude, so in quantum mechanics the task is to find an equation governing the quantity ψ rather than the probability density. For this equation to correspond to the similarity of the diffraction patterns observed for electrons and electromagnetic waves, it must, like Maxwell's equations, be linear.

Indeed, for the amplitudes to cancel out mutually they must be added up at every point in space. But for the sum of two solutions of the equation to satisfy it again, that is, also be its solution, the equation must necessarily be linear. Therefore, the equation governing the wave function is linear. From such an equation can be determined the wave phase, without which it is impossible to construct the diffraction pattern.

We thus obtain one of the prime principles of quantum mechanics: the sum of the two solutions of the wave equation is also its solution. This assertion is known as the *superposition principle*. Together with the uncertainty principle, which states the condition for the transformation of quantum equations into the classical equations of motion, it provides the approach to the fundamental equation of quantum mechanics.

It is convenient to start by writing not the equation itself, but its solution, that is, the wave function, as applied to a free particle. Such a particle possesses an exactly defined and constant momentum **p**. From the de Broglie relation (22.1), this momentum corresponds to the vector of a certain wave, $\mathbf{k} = \mathbf{p}/h$. Indeed, if electrons with momentum **p** are beamed on a crystal, they will present the same diffraction pattern as a wave with the wave vector $\mathbf{k} = \mathbf{p}/h$. It follows that the wave function of a free particle depends on the coordinates in the following way:

$$\psi \propto e^{i\mathbf{kr}} = e^{i\mathbf{pr}/h}$$

Only the square of the modulus of the wave function can be measured, but not the function itself, which is why it is written in complex form. For physical considerations, the wave function cannot be required to be real.

The dependence of the wave function on time is also easily determined if we recall that wave frequency corresponds to the energy of a particle in the same sense that the wave vector corresponds to the momentum. It was shown in Section 21 that the coefficient of proportionality between ω and E should be the same as between **k** and **p**. Therefore

$$\omega = E/h \tag{23.2}$$

From this we obtain the expression for the wave function of a free particle:

$$\psi = e^{-i\omega t + i\mathbf{kr}} = e^{-iEt/h + i\mathbf{pr}/h} \tag{23.3}$$

The group velocity of the waves is

$$\mathbf{v} = \frac{\partial \omega}{\partial \mathbf{k}} = \frac{\partial E}{\partial \mathbf{p}} \tag{23.4}$$

Thus, it coincides with the velocity of the particles, as it should in accordance with Section 21.

But it is now apparent that the wave function (23.3) is simply related to the action of a free particle:

$$\psi = e^{iS/h} \tag{23.5}$$

with S the action. Indeed, it is in this case equal to

$$S = -Et + \mathbf{pr} \tag{23.6}$$

in agreement with the fact that

$$\mathbf{p} = \frac{\partial S}{\partial \mathbf{r}} \equiv \text{grad } S, \qquad E = -\frac{\partial S}{\partial t}$$

as it should be according to the equations of Section 21, which establish the analogy between mechanics and optics. Equation (23.5) confirms the relationship between the phase of a wave and the action of a particle obtained in that section.

The Wave Equation for a Free Particle. We shall now develop the wave equation for the just obtained function ψ. Naturally enough, the form of this equation is related to the fact that E is expressed in terms of $\mathbf{p}$, or what is the same thing, ω in terms of $\mathbf{k}$. For example, in optics, where $\omega^2 = c^2 k^2$, the wave equation should be written as

$$\nabla^2 \varphi - \frac{1}{c^2} \frac{\partial^2 \varphi}{\partial t^2} = 0$$

It is relativistically invariant. Our purpose is to first develop an equation of nonrelativistic mechanics, in which $E = p^2/(2m)$. As we know, the analogy between optics and mechanics in no way requires that the latter's equations be written in a relativistically invariant form. The meaning of the analogy consists in the correspondence of the quantities. This is sufficient to obtain the equation for the wave function (23.3). We have

$$\frac{\partial \psi}{\partial t} = -\frac{iE}{h} \psi \tag{23.7}$$

$$\frac{\partial \psi}{\partial x} = \frac{ip_x}{h} \psi, \qquad \frac{\partial^2 \psi}{\partial x^2} = -\frac{p_x^2}{h^2} \psi \tag{23.8}$$

From (23.7) and (23.8) we obtain

$$-\frac{h}{i} \frac{\partial \psi}{\partial t} = E\psi = -\frac{h^2}{2m} \left(\frac{\partial^2 \psi}{\partial x^2} + \frac{\partial^2 \psi}{\partial y^2} + \frac{\partial^2 \psi}{\partial z^2} \right) = \frac{p^2}{2m} \psi \tag{23.9}$$

that is, a nonrelativistic equation for the wave function, ψ. In terms of the Laplace operator it is expressed as follows:

$$-\frac{h}{i} \frac{\partial \psi}{\partial t} = -\frac{h^2}{2m} \nabla^2 \psi \tag{23.10}$$

Equation (23.9) is valid because $E = p^2/(2m)$.

The Schrödinger Equation. Let us generalize Eq. (23.10) for the case of a particle moving in an external potential field $U(\mathbf{r})$. To obtain a relationship $E = p^2/(2m) + U$ analogous to $E = p^2/(2m)$ for a free particle, we must put

$$-\frac{h}{i} \frac{\partial \psi}{\partial t} = -\frac{h^2}{2m} \nabla^2 \psi + U\psi \tag{23.11}$$

This equation was developed in 1926 by Erwin Schrödinger, who generalized the de Broglie relations for the case of bound electrons.

Equation (23.11) follows directly from (23.10) for the simplest case of $U =$ constant, because then it is satisfied by the same substitution of (23.3) but for the value of momentum $p = [2m (E - U)]^{1/2}$. From here it is but one step to the generalization for the case of variable potential energy.

This generalization, however, can in no circumstances be regarded as the "derivation" of a quantum mechanical equation from principles or equations of pre-quantum, classical physics. The Schrödinger equation expresses a new physical law.

The authenticity of Eq. (23.11) is seen in the limiting transition to classical physics, similar to the transition from wave to geometrical optics.

Let us assume that Eq. (23.5) involves the action not of a free particle but of one moving in a field $U(\mathbf{r})$, and find the approximation that satisfies the Schrödinger equation (23.11), if the wave function in it is expressed in terms of the action.

We determine the values of the derivatives of the wave function:

$$\frac{\partial \psi}{\partial t} = \frac{\partial}{\partial t} e^{iS/h} = \frac{i}{h} \frac{\partial S}{\partial t} \psi$$

$$\frac{\partial \psi}{\partial x} = \frac{\partial}{\partial x} e^{iS/h} = \frac{i}{h} \frac{\partial S}{\partial x} \psi$$

$$\frac{\partial^2 \psi}{\partial x^2} = \frac{i}{h} \frac{\partial^2 S}{\partial x^2} \psi - \frac{1}{h^2} \left(\frac{\partial S}{\partial x} \right)^2 \psi$$

substitute them into Eq. (23.11) and cancel out h^{-2}. This leaves the following equation for S:

$$-\frac{\partial S}{\partial t} = \frac{1}{2m} \left[\left(\frac{\partial S}{\partial x} \right)^2 + \left(\frac{\partial S}{\partial y} \right)^2 + \left(\frac{\partial S}{\partial z} \right)^2 \right] - \frac{ih}{2m} \nabla^2 S + U \tag{23.12}$$

It is best to pass to the limit of classical mechanics by assuming that the action quantum h tends to zero. This is similar to the assertion that all classical quantities characterizing motion are large in comparison with h. Then from (23.12) we obtain

$$-\frac{\partial S}{\partial t} = \frac{(\operatorname{grad} S)^2}{2m} + U \tag{23.13}$$

which is the Hamilton-Jacobi equation.

The limiting process performed here almost exactly repeats the transition from wave to geometrical optics carried out in Section 21. Indeed, if we put $h = 0$, this is equivalent to the vanishing of the de Broglie wavelength, which corresponds to a transition from waves to paths.

19—0452

We thus see that the Schrödinger wave equation does in fact yield a correct limiting process. It is as it were the fourth term in the relationship

$$\text{geometrical optics} \rightarrow \text{classical mechanics}$$
$$\downarrow \qquad\qquad\qquad\qquad \downarrow$$
$$\text{wave optics} \longrightarrow \boxed{\text{quantum mechanics}}$$
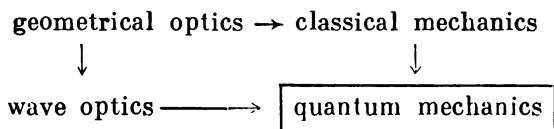
The vertical arrows denote the passage from rays or paths to the wave picture; the horizontal arrows denote the passage from waves to particles. The latter refers only to nonquantized electromagnetic field equations, since quantization requires corpuscular representation of light quanta. The analogy here is between quantum mechanics and classical wave optics.

The Range of Applicability of Various Theories. The regions in which quantum mechanics and wave optics can be applied do not, strictly speaking, overlap anywhere; in wave optics or, what is just the same, electrodynamics, the velocity of light c is considered finite, and the quantum of action h is considered arbitrarily small. In nonrelativistic quantum mechanics c is considered arbitrarily large, while h has a finite value. A quantum theory of the electromagnetic field, in which both h and c have finite values (that is, the velocity ranges are comparable with c, and quantities with the dimensions of action are comparable with h), has, in essentials, also been completed. At any rate, any concrete problem requiring the application of quantum electrodynamics, may be uniquely solved to any required degree of precision, and the results agree with experiment.

Nonrelativistic quantum mechanics based on the relationship $E = p^2/(2m) + U$ is, within its sphere of application, as complete a theory as Newtonian mechanics. Like the equations of Newtonian mechanics, the wave equation (23.11) is valid only for particle velocities that are sufficiently small in comparison with the speed of light, but in the sphere of its application it is as firmly established as Newton's laws of motion are for macroscopic bodies.

Of course, quantum mechanics continues to perfect its methods of approach to various specific problems. The basis for this is the correctness of its general propositions. In this sense, Newtonian mechanics, too, is continuing to advance to this day.

The Normalization Condition for a Wave Function. Let us return to the wave equation (23.11). We write it for a wave function ψ and the complex conjugate ψ^*, in the equation for which we must re-

place i by $-i$:

$$-\frac{h}{i}\frac{\partial\psi}{\partial t} = -\frac{h^2}{2m}\nabla^2\psi + U\psi$$

$$\frac{h}{i}\frac{\partial\psi^*}{\partial t} = -\frac{h^2}{2m}\nabla^2\psi^* + U\psi^*$$

We multiply the first equation by ψ^*, and the second by ψ, and subtract the second from the first. The term $\psi^* U\psi$ is eliminated and the remaining terms give

$$-\frac{h}{i}\psi^*\frac{\partial\psi}{\partial t} - \frac{h}{i}\psi\frac{\partial\psi^*}{\partial t} = -\frac{h^2}{2m}(\psi^*\nabla^2\psi - \psi\nabla^2\psi^*) \qquad (23.14)$$

The left-hand side of the last equality is transformed to the form

$$-\frac{h}{i}\frac{\partial}{\partial t}\psi^*\psi = -\frac{h}{i}\frac{\partial}{\partial t}|\psi|^2$$

We can write the right-hand side explicitly thus:

$$-\frac{h^2}{2m}(\psi^*\nabla^2\psi - \psi\nabla^2\psi^*) = -\frac{h^2}{2m}(\psi^*\,\mathrm{div}\,\mathrm{grad}\,\psi$$
$$-\psi\,\mathrm{div}\,\mathrm{grad}\,\psi^*)$$

$$= -\frac{h^2}{2m}\,\mathrm{div}\,(\psi^*\,\mathrm{grad}\,\psi - \psi\,\mathrm{grad}\,\psi^*)$$

(see (11.27)). Finally, we represent the equality in the following form:

$$\frac{\partial}{\partial t}|\psi|^2 = -\mathrm{div}\left[\frac{h}{2mi}(\psi^*\,\mathrm{grad}\,\psi - \psi\,\mathrm{grad}\,\psi^*)\right] \qquad (23.15)$$

The left-hand side of this equality is the time derivative of the probability density of finding a particle close to some point of space. Let us integrate (23.15) over the whole volume in which the particle might be situated. If this volume is finite, then beyond its boundaries ψ and ψ^* must be equal to zero. But then, from the Gauss theorem, the right-hand side of (23.15), transformed into a surface integral (the surface being outside the volume), vanishes:

$$\frac{\partial}{\partial t}\int|\psi|^2\,dV = 0 \qquad (23.16)$$

It follows that the integral itself does not depend upon time. If it has a definite value, as may be expected in integration over a finite volume, it can be given a simple physical meaning, namely that it must be equal to the probability of an electron, or any particle in general, occurring somewhere within a volume in which it is bound to be. Such a probability is equal to certainty, or unity. Thus

$$\int|\psi|^2\,dV = 1 \qquad (23.17)$$

This equation is called the *normalization condition* for a wave function.

If the motion is infinite, the integral involved in (23.17) is not finite. In that case we can either vary the normalization condition (see Sec. 25) or normalize the probability to a larger, but finite, volume and then tend it to infinity. The physical results are, of course, not affected by this.

If we integrate (23.15) over an arbitrary volume, we obtain

$$\frac{\partial}{\partial t} \int |\psi|^2 \, dV = - \int \mathrm{div} \left[\frac{h}{2mi} \left(\psi^* \, \mathrm{grad} \, \psi - \psi \, \mathrm{grad} \, \psi^* \right) \right] dV$$

$$= - \int \frac{h}{2mi} \left(\psi^* \, \mathrm{grad} \, \psi - \psi \, \mathrm{grad} \, \psi^* \right) dS \quad (23.18)$$

On the left we have the probability of an electron occurring within the given volume, on the right is the probability flux across the boundary surface of the volume. It is apparent from this that, according to (23.18), the density of the probability flux is

$$\mathbf{j} = \frac{h}{2mi} \left(\psi^* \, \mathrm{grad} \, \psi - \psi \, \mathrm{grad} \, \psi^* \right) \quad (23.19)$$

It follows that a real function yields $\mathbf{j} = 0$ and does not describe electron current. Therefore the general definition of a wave function must of necessity involve complex quantities.

The Equation for Stationary States. Suppose that the potential energy does not depend on time explicitly. Then in classical mechanics, as we know, the law of conservation of energy of the system holds. The action of such a system involves the term $-Et$. But since in the classical limit $\psi = e^{iS/h}$, we shall seek ψ proportional to $e^{-iEt/h}$ in the most general case as well:

$$\psi = e^{-iEt/h} \psi_0 (x, y, z) \quad (23.20)$$

Substituting this expression into (23.11) and abandoning the zero subscript, we obtain the equation

$$-\frac{h^2}{2m} \nabla^2 \psi + U \psi = E \psi \quad (23.21)$$

In this equation the energy E should be treated as a certain constant quantity. If we are then concerned with the states corresponding to finite motions, the probability of a particle occurring at an infinite distance from the domain in which ψ is other than zero must be infinitesimal. Analyses of specific examples reveal that the latter holds not for all values of the energy E, but only for certain values belonging to a discrete, limited assembly.

If, for example, $U(\infty) = 0$, for negative values of E we have two values of the function $\psi \sim \exp \left[\pm (2m| E |)^{1/2} r/h \right]$. One of

them increases exponentially as $r \to \infty$, and there is no way of normalizing it, because at infinite distance from the origin of the coordinate system $\exp \left[(2m \mid E \mid)^{1/2} r/h \right]$ is infinite. There are, however, select values of E for which the coefficient of the exponentially increasing solution vanishes. It follows that in finite motion only certain definite values of E are possible. Their totality is called the *energy spectrum* of the system.

The state of a system corresponding to such an energy value is called *stationary*, and the value itself is known as the *energy eigenvalue*. In general, any specific value of energy that corresponds to a definite wave function in Eq. (23.21) which, in turn, satisfies appropriate boundary conditions is called the energy eigenvalue. The choice of boundary conditions depends on the type of motion being investigated.

Thus, we find that, unlike energy in classical mechanics, in quantum mechanics energy cannot be stated arbitrarily.

24

OPERATORS IN QUANTUM MECHANICS

Momentum Eigenvalues. At the end of the preceding section we defined energy eigenvalues on the basis of the wave equation (23.21). However, it is also very important to find the eigenvalues of other quantities: linear momentum, angular momentum, etc. In order to obtain the respective equations it is convenient to proceed from the form of ψ in passing to the limit of classical mechanics:

$$\psi = e^{iS/h} \tag{24.1}$$

as it was done in deriving Eq. (23.21).

Let us apply the operation $(h/i)(\partial/\partial x)$ to both sides of Eq. (24.1), that is, we take the partial derivative with respect to x and multiply by h/i to get

$$\frac{h}{i} \frac{\partial \psi}{\partial x} = \frac{\partial S}{\partial x} e^{iS/h} = \frac{\partial S}{\partial x} \psi \tag{24.2}$$

But in the classical limit S becomes the action of the particle, while $\partial S/\partial x$ becomes the component of momentum p_x (see (10.24)). Therefore, the eigenvalue equation for momentum that yields the correct transition to classical mechanics is of the form

$$\frac{h}{i} \frac{\partial \psi}{\partial x} = p_x \psi \tag{24.3}$$

where p_x is the eigenvalue of the momentum projection on the x axis.

Momentum and Energy Operators. Let us compare Eq. (24.3) with the wave equation (23.21):

$$\frac{1}{2m}\left[\left(\frac{h}{i}\frac{\partial}{\partial x}\right)^2+\left(\frac{h}{i}\frac{\partial}{\partial y}\right)^2+\left(\frac{h}{i}\frac{\partial}{\partial z}\right)^2\right]\psi+U\psi=E\psi \quad (24.4)$$

Here, the symbol $(\partial/\partial x)^2$ denotes the second derivative, which must then be applied to ψ.

In order to find the energy and momentum eigenvalues we must perform a definite set of differential operations and multiplications by the function of coordinates in the left part of Eqs. (23.21) and (24.3). But these sets are connected in a very curious manner, as will now be shown. We shall call the symbol $\partial/\partial x$ multiplied by h/i the *momentum operator* applied to a wave function. Instead of $(h/i)(\partial/\partial x)$ we symbolically write $\hat{p}_x$. Then it is necessary to rewrite Eq. (24.3) as

$$\hat{p}_x\psi=p_x\psi \tag{24.5}$$

This equation denotes exactly the same as (24.3), but the symbolic notation $\hat{p}_x$ emphasizes that the corresponding operation is applied in order to find the momentum eigenvalues.

The operation on the left-hand side of (24.4) we shall also symbolically denote by $\hat{\mathscr{H}}$. We write $\hat{\mathscr{H}}$ and not $\hat{E}$ because the energy is assumed to be expressed in terms of momentum, similar to the Hamiltonian $\mathscr{H}$. Then, in shorter notation, (24.4) appears as

$$\hat{\mathscr{H}}\psi=E\psi \tag{24.6}$$

Comparing (24.4) and (24.3), we see that the momentum and energy operators are related by the same equations as the corresponding quantities:

$$\hat{\mathscr{H}}=\frac{1}{2m}(\hat{p}_x^2+\hat{p}_y^2+\hat{p}_z^2)+\hat{U} \tag{24.7}$$

We have written $\hat{U}$ instead of simply U to emphasize that in this equation $\hat{U}$ is regarded not as an independent quantity but as an operator operating on ψ, that is, a multiplication operator of $U(\mathbf{r})$ by ψ. Equation (24.7) is symbolic—it is understood that both sides are applied to ψ.

The usefulness of abbreviated operator notation in quantum mechanics is that the equations thus become more expressive. The relation between quantum laws of motion and classical laws, which are limiting cases with respect to the quantum ones, can be best of all seen in operator notation.

If in classical equations relating mechanical quantities we replace the momenta by their operators, we then obtain correct operator relationships of quantum mechanics. The limiting transition to

classical mechanics restores the usual relationships between quantities. Indeed, in the limiting transition (24.1), the operator $\hat{\mathbf{p}} = (h/i)\nabla$ gives $\mathbf{p}\psi$. If we must perform the limiting transition for $\hat{\mathbf{p}}^2$, then we need to differentiate only the exponential each time, because this yields the quantum of action in the denominator. For $h \to 0$, only terms with the highest degree of h in the denominator remain, and it is these very terms that are obtained in replacing the operator $\hat{\mathbf{p}}$ by the quantity grad S (that is, by the classical momentum vector).

But the advantages of operator symbolism in quantum mechanics are not restricted to limiting processes. They will be made apparent in the following discourse.

The Operator of an Angular Momentum Component. It is now easy to determine the operator of the angular momentum projection M_z. We know from Section 5 that the angular momentum M_z is at the same time the generalized momentum corresponding, like the generalized coordinate, to the angle of rotation about the z axis: $M_z = p_\varphi$. Then, from (10.24)

$$p_\varphi = \frac{\partial S}{\partial \varphi} \tag{24.8}$$

It is therefore clear why in quantum mechanics the operator $\hat{p}_\varphi$ should have the form

$$\hat{p}_\varphi = \frac{h}{i} \frac{\partial}{\partial \varphi} \tag{24.9}$$

But according to classical mechanics the projection M_z is related to the Cartesian projections of linear momentum as follows:

$$M_z = xp_y - yp_x \tag{24.10}$$

Hence, there should exist the operator relationship

$$\hat{p}_\varphi = \hat{M}_z = \hat{x}\hat{p}_y - \hat{y}\hat{p}_x = \frac{h}{i}\left(x\frac{\partial}{\partial y} - y\frac{\partial}{\partial x}\right) \tag{24.11}$$

Let us verify that the definitions (24.9) and (24.11) do indeed coincide. We pass to cylindrical coordinates

$$x = r\cos\varphi, \qquad y = r\sin\varphi$$

whence we have

$$\frac{\partial\psi}{\partial x} = \frac{\partial\psi}{\partial r}\frac{\partial r}{\partial x} + \frac{\partial\psi}{\partial\varphi}\frac{\partial\varphi}{\partial x}, \qquad \frac{\partial\psi}{\partial y} = \frac{\partial\psi}{\partial r}\frac{\partial r}{\partial y} + \frac{\partial\psi}{\partial\varphi}\frac{\partial\varphi}{\partial y}$$

Expressing cylindrical coordinates in terms of Cartesian, we write

$$r = (x^2 + y^2)^{1/2}, \qquad \frac{\partial r}{\partial x} = \frac{x}{r} = \cos\varphi, \qquad \frac{\partial r}{\partial y} = \frac{y}{r} = \sin\varphi$$

$$\varphi = \text{arc cot}\,\frac{y}{x}, \qquad \frac{\partial\varphi}{\partial x} = -\frac{y}{r^2} = -\frac{\sin\varphi}{r}, \qquad \frac{\partial\varphi}{\partial y} = \frac{x}{r^2} = \frac{\cos\varphi}{r}$$

Substituting all these expressions into (24.11), we find that it is indeed identical in meaning to (24.9):

$$\frac{h}{i}\left(x\,\frac{\partial\psi}{\partial y}-y\,\frac{\partial\psi}{\partial x}\right)=\frac{h}{i}\left[r\cos\varphi\left(\sin\varphi\,\frac{\partial\psi}{\partial r}+\frac{\cos\varphi}{r}\,\frac{\partial\psi}{\partial\varphi}\right)\right.$$

$$\left.-r\sin\varphi\left(\cos\varphi\,\frac{\partial\psi}{\partial r}-\frac{\sin\varphi}{r}\,\frac{\partial\psi}{\partial\varphi}\right)\right]$$

$$=\frac{h}{i}\left(\cos^2\varphi+\sin^2\varphi\right)\frac{\partial\psi}{\partial\varphi}=\frac{h}{i}\,\frac{\partial\psi}{\partial\varphi}$$

The other two angular momentum projections can be determined similarly to (24.10), but first it is necessary to establish the quantity that can exist simultaneously in the same physical state of the system. We saw in Section 22, for example, that the coordinate and corresponding linear momentum projection cannot exist simultaneously. Let us now develop a general criterion in terms of the operators of the corresponding quantities.

The Simultaneous Existence of Two Physical Quantities. Suppose that in a certain state described by the wave function ψ there simultaneously exist two physical quantities, λ and ν. Let us find the necessary condition for this.

If in a certain state it is possible to define a physical quantity λ, the wave function of that state must be the eigenfunction of the operator $\hat{\lambda}$. If a quantity ν is also defined in that state, then the same function ψ satisfies two equations

$$\hat{\lambda}\psi=\lambda\psi \tag{24.12}$$

$$\hat{\nu}\psi=\nu\psi \tag{24.13}$$

In other words, the function ψ is the eigenfunction of both the operator $\hat{\lambda}$ and the operator $\hat{\nu}$. Let us now operate with $\hat{\nu}$ on (24.12), remembering that in the left-hand side of this equation we have simply the quantity λ, not an operator; we similarly operate with $\hat{\lambda}$ on (24.13). We then obtain

$$\hat{\nu}\hat{\lambda}\psi=\hat{\nu}\lambda\psi=\lambda\hat{\nu}\psi=\lambda\nu\psi \tag{24.14a}$$

$$\hat{\lambda}\hat{\nu}\psi=\hat{\lambda}\nu\psi=\nu\hat{\lambda}\psi=\nu\lambda\psi \tag{24.14b}$$

Now subtract (24.14b) from (24.14a) to get

$$\hat{\nu}\,\hat{\lambda}\psi-\hat{\lambda}\,\hat{\nu}\psi=\lambda\nu\psi-\nu\lambda\psi=0$$

Since only the derivatives of numbers appear in the right-hand side, they cancel out, and we arrive at the required condition:

$$\hat{\nu}\hat{\lambda}\psi=\hat{\lambda}\hat{\nu}\psi \tag{24.15}$$

Equation (24.15) can be written symbolically as an equality between operators:

$$\hat{v}\hat{\lambda} = \hat{\lambda}\hat{v}, \qquad \hat{v}\hat{\lambda} - \hat{\lambda}\hat{v} = 0 \tag{24.16}$$

This symbolic equality means that the result of a successive operation with $\hat{\lambda}$ and $\hat{v}$ should not depend on the order of operation, otherwise Eqs. (24.12) and (24.13) cannot have the common eigenfunction ψ, that is, a common solution.

We have proved here the necessity of condition (24.16) for the quantities λ and ν to exist simultaneously in the same state of a system to which the wave function ψ corresponds. We could also show that this condition is sufficient, but we shall not go into the proof here.

Commutation Relations for Certain Operators. Let us now apply the obtained result to two quantities which definitely do not exist in the same state, the coordinate x and momentum p_x. We must calculate the commutator $\hat{p}_x\hat{x} - \hat{x}\hat{p}_x$.

Changing from symbolic notation to the usual one, we obtain

$$\frac{h}{i}\frac{\partial}{\partial x}x\psi - x\frac{h}{i}\frac{\partial}{\partial x}\psi = \frac{h}{i}\psi + x\frac{h}{i}\frac{\partial\psi}{\partial x} - x\frac{h}{i}\frac{\partial\psi}{\partial x} = \frac{h}{i}\psi$$

Reverting to symbolic notation, we represent the obtained result in the following form:

$$\hat{p}_x\hat{x} - \hat{x}\hat{p}_x = \frac{h}{i} \tag{24.17}$$

Thus, the result of operating with $\hat{p}_x$ and $\hat{x}$ depends upon the order of their action; $\hat{p}_x$ and $\hat{x}$ do not commute. And this was to be expected because the quantities p_x and x do not exist simultaneously.

The eigenfunction of the operator $\hat{x}$ satisfies the equation $(\hat{x} - x')\psi = 0$. Consequently, it is equal to zero over the whole region where the coordinate x is not equal to the chosen eigenvalue x'. This function differs from zero only at one point $x = x'$. The eigenfunction of the momentum operator which satisfies Eq. (24.3) is $e^{ip_x x/h}$; it differs from zero over all the space. This example shows how great the difference is between the eigenfunctions of operators that do not commute. The abbreviated notation (24.17) is a convenient representation of the preceding equation in which the momentum operator is expressed explicitly in terms of the derivative. In future we shall write commutation relations exclusively in operator form, because the symbolic notation is more concise and vivid: one immediately sees the quantities represented by the operators.

The various momentum components commute, since

$$\hat{p}_x\hat{p}_y - \hat{p}_y\hat{p}_x = 0 \tag{24.18}$$

(the word "operator" will be frequently omitted in future as being self-evident). The commutation relation (24.18) is obtained simply from the fact that the result of applying two partial derivatives does not depend upon the order of differentiation. It is also obvious that

$$\hat{p}_y \hat{x} - \hat{x} \hat{p}_y = 0 \tag{24.19}$$

In tensor notation these commutation relations take the form

$$\hat{p}_\alpha \hat{x}_\beta - \hat{x}_\beta \hat{p}_\alpha = \frac{h}{i} \delta_{\alpha\beta} \tag{24.20}$$

$$\hat{p}_\alpha \hat{p}_\beta - \hat{p}_\beta \hat{p}_\alpha = 0 \tag{24.21}$$

$$\hat{x}_\alpha \hat{x}_\beta - \hat{x}_\beta \hat{x}_\alpha = 0 \tag{24.22}$$

We now calculate the commutator of two angular momentum components, $\hat{M}_x$ and $\hat{M}_y$, where

$$\hat{M}_x = \hat{y} \hat{p}_z - \hat{z} \hat{p}_y, \qquad \hat{M}_y = \hat{z} \hat{p}_x - \hat{x} \hat{p}_z$$

We group the terms without violating the order of the coordinates and corresponding linear momenta to get

$$\hat{M}_x \hat{M}_y - \hat{M}_y \hat{M}_x = (\hat{y} \hat{p}_z - \hat{z} \hat{p}_y)(\hat{z} \hat{p}_x - \hat{x} \hat{p}_z)$$
$$- (\hat{z} \hat{p}_x - \hat{x} \hat{p}_z)(\hat{y} \hat{p}_z - \hat{z} \hat{p}_y)$$

Making use of the commutation relation $\hat{p}_z \hat{z} - \hat{z} \hat{p}_z = h/i$, we obtain the required result:

$$\hat{M}_x \hat{M}_y - \hat{M}_y \hat{M}_x = \hat{y} \hat{p}_x (\hat{p}_z \hat{z} - \hat{z} \hat{p}_z) - \hat{x} \hat{p}_y (\hat{p}_z \hat{z} - \hat{z} \hat{p}_z)$$
$$= \frac{h}{i} (\hat{y} \hat{p}_x - \hat{x} \hat{p}_y) = ih\hat{M}_z \tag{24.23a}$$

Changing the indices x, y, z cyclically, we obtain the remaining commutation relations:

$$\hat{M}_y \hat{M}_z - \hat{M}_z \hat{M}_y = ih\hat{M}_x \tag{24.23b}$$

$$\hat{M}_z \hat{M}_x - \hat{M}_x \hat{M}_z = ih\hat{M}_y \tag{24.23c}$$

All three commutation relations can be easily remembered if we write them in contracted vector form as

$$\hat{\mathbf{M}} \times \hat{\mathbf{M}} = ih\hat{\mathbf{M}} \tag{24.24}$$

Expanding this equality in components, we once again arrive at (24.23a)-(24.23c).

It will be noted that the vector product of an operator by itself may not equal zero if the operator components of the vector do not commute (but, for example, $\hat{\mathbf{p}} \times \hat{\mathbf{p}} = 0$).

The Square of the Angular Momentum. Let us now examine further properties of angular momentum. We shall show that even though two angular momentum projections do not exist, a single angular momentum projection exists together with its square

$$\hat{M}^2 = \hat{M}_x^2 + \hat{M}_y^2 + \hat{M}_z^2 \tag{24.25}$$

Let us verify this. For this we find the commutator

$$\hat{M}^2 \hat{M}_z - \hat{M}_z \hat{M}^2 = \hat{M}_x^2 \hat{M}_z - \hat{M}_z \hat{M}_x^2 + \hat{M}_y^2 \hat{M}_z - \hat{M}_z \hat{M}_y^2$$

where the equality holds because M_z and M_z^2, of course, commute. Let us add to the right-hand side of the last equality and subtract from it the combinations $\hat{M}_x \hat{M}_z \hat{M}_x$ and $\hat{M}_y \hat{M}_z \hat{M}_y$; we take $\hat{M}_x$ and $\hat{M}_y$ outside the brackets, once on the right and once on the left. Then we obtain

$$\hat{M}^2 \hat{M}_z - \hat{M}_z \hat{M}^2 = \hat{M}_x (\hat{M}_x \hat{M}_z - \hat{M}_z \hat{M}_x)$$
$$+ (\hat{M}_x \hat{M}_z - \hat{M}_z \hat{M}_x) \hat{M}_x + \hat{M}_y (\hat{M}_y \hat{M}_z - \hat{M}_z \hat{M}_y)$$
$$+ (\hat{M}_y \hat{M}_z - \hat{M}_z \hat{M}_y) \hat{M}_y$$
$$= - ih\hat{M}_x \hat{M}_y - ih\hat{M}_y \hat{M}_x + ih\hat{M}_y \hat{M}_x$$
$$+ ih\hat{M}_x \hat{M}_y = 0 \quad (24.26)$$

Here we have made use of the commutation relations for separate angular momentum components.

For subsequent applications the operator of the angular momentum square must be transformed to spherical coordinates. For that it is useful first to go over to tensor notation. For a separate component we have

$$\hat{M}_\alpha = \varepsilon_{\alpha\beta\gamma} \hat{x}_\beta \hat{p}_\gamma \tag{24.27}$$

whence the operator of the square of the angular momentum is

$$\hat{M}^2 = \hat{M}_\alpha \hat{M}_\alpha = \varepsilon_{\alpha\beta\gamma} \varepsilon_{\alpha\zeta\eta} \hat{x}_\beta \hat{p}_\gamma \hat{x}_\zeta \hat{p}_\eta$$

Now we take advantage of the fact that $\varepsilon_{\alpha\beta\gamma} \varepsilon_{\alpha\zeta\eta} = \delta_{\beta\zeta} \delta_{\gamma\eta} - \delta_{\beta\eta} \delta_{\gamma\zeta}$ which is proved by comparison with the vector equation $(\mathbf{A} \times \mathbf{B})(\mathbf{C} \times \mathbf{D}) = (\mathbf{A} \cdot \mathbf{C})(\mathbf{B} \cdot \mathbf{D}) - (\mathbf{A} \cdot \mathbf{D})(\mathbf{B} \cdot \mathbf{C})$. Then $\hat{M}^2$ can be rewritten as

$$\hat{M}^2 = \hat{x}_\beta \hat{p}_\gamma \hat{x}_\beta \hat{p}_\gamma - \hat{x}_\beta \hat{p}_\gamma \hat{x}_\gamma \hat{p}_\beta$$

After performing certain manipulations according to the rules (24.20)-(24.22), we reduce the right-hand side of the obtained expres-

sion to

$$\hat{M}^2 = \frac{h}{i}\,\hat{x}_\alpha \delta_{\alpha\beta}\hat{p}_\beta + \hat{x}_\alpha^2\hat{p}_\beta^2 - \frac{h}{i}\,\delta_{\alpha\beta}\hat{x}_\alpha\hat{p}_\beta - \frac{h}{i}\,\hat{x}_\alpha\hat{p}_\alpha - (\hat{x}_\alpha\hat{p}_\alpha)^2$$

$$= \hat{x}_\alpha^2\hat{p}_\beta^2 - \frac{h}{i}\,\hat{x}_\alpha\hat{p}_\alpha - (\hat{x}_\alpha\hat{p}_\alpha)^2$$

But the expression $\hat{x}_\alpha\hat{p}_\alpha$ is just $(h/i)\,(\hat{r}\nabla)$, which in spherical coordinates is rewritten simply as $(h/i)r(\partial/\partial r)$. Furthermore, $\hat{p}_\beta^2 = -h^2\nabla^2$. Now substitute the Laplacian, written in terms of the spherical coordinates, according to (11.46). We then see that the square of the angular momentum is expressed simply in terms of the angular part of the Laplacian:

$$\hat{M}^2 = -h^2 r^2 \left(\frac{1}{r^2}\frac{\partial}{\partial r}r^2\frac{\partial}{\partial r} + \frac{1}{r^2\sin\vartheta}\frac{\partial}{\partial\vartheta}\sin\vartheta\frac{\partial}{\partial\vartheta} \right.$$
$$\left. + \frac{1}{r^2\sin^2\vartheta}\frac{\partial^2}{\partial\varphi^2} \right) + h^2 r\frac{\partial}{\partial r} + h^2 r\frac{\partial}{\partial r}r\frac{\partial}{\partial r}$$

$$= -h^2 \left(\frac{1}{\sin\vartheta}\frac{\partial}{\partial\vartheta}\sin\vartheta\frac{\partial}{\partial\vartheta} + \frac{1}{\sin^2\vartheta}\frac{\partial^2}{\partial\varphi^2} \right) \qquad (24.28)$$

The Eigenfunctions of the Angular Momentum Square and the Angular Momentum Projection. Let us now find the eigenfunction of the square of the angular momentum, that is, the solution of the equation

$$\hat{M}^2\psi = M^2\psi \qquad (24.29)$$

For this we proceed from the well-known equality

$$\nabla^2\frac{1}{r} = 0 \qquad (24.30)$$

We take l arbitrary constant vectors $\mathbf{a}_1$, $\mathbf{a}_2$, ..., $\mathbf{a}_i$, ..., $\mathbf{a}_l$ and form the following set of operators: $(\mathbf{a}_1\cdot\nabla)$, $(\mathbf{a}_2\cdot\nabla)$, ..., $(\mathbf{a}_i\cdot\nabla)$, ..., $(\mathbf{a}_l\cdot\nabla)$. Since the vectors are constant, all of the written operators commute with the Laplacian. After applying all the operators, we therefore write (24.30) in the form

$$\nabla^2 (\mathbf{a}_1\cdot\nabla)(\mathbf{a}_2\cdot\nabla)\ldots(\mathbf{a}_i\cdot\nabla)\ldots(\mathbf{a}_l\cdot\nabla)\frac{1}{r} = 0 \qquad (24.31)$$

We express the Laplacian in terms of spherical coordinates and make use of (24.28) to get

$$\nabla^2 = \frac{1}{r^2}\frac{\partial}{\partial r}r^2\frac{\partial}{\partial r} - \frac{\hat{M}^2}{h^2 r^2} \qquad (24.32)$$

Now we take into account the circumstance that every operation, with $(\mathbf{a}_i\cdot\nabla)$ on r^{-1} raises the power of r in the denominator by unity, while in the numerator there is always found some function of the

angles. Therefore the result of operating l times with $(\mathbf{a}_i \cdot \nabla)$ on r^{-1} can be represented as

$$(\mathbf{a}_1 \cdot \nabla) \ldots (\mathbf{a}_l \cdot \nabla) \frac{1}{r} = \frac{f(\vartheta, \varphi)}{r^{l+1}} \tag{24.33}$$

We transfer the part of the Laplacian involving differentiation with respect to r to the right-hand side of the equation. Then, differentiating and cancelling out $r^{-(l+3)}$, we arrive at the equation

$$\hat{M}^2 f(\vartheta, \varphi) = h^2 l \, (l+1) \, f(\vartheta, \varphi) \tag{24.34}$$

But this equation coincides in form with the eigenvalue equation for the operator $\hat{M}^2$ (24.29). Hence, we have found the eigenfunctions of the operator $\hat{M}^2$ and its eigenvalues $h^2 l \, (l+1)$.

The fact that the square of the quantity is equal to $h^2 l \, (l+1)$ and not to the total square may cause some surprise. It would appear that if vector **M** is directed along a coordinate axis, the eigenvalue of the square should be equal to the square of the eigenvalue of the projection (we shall soon see that it is an integer). Actually, though, the angular momentum projections do not commute, so that if one of them takes a certain given value the others do not have any definite values, zero included. For this reason the expression for the angular momentum square is more involved than simply the square of an integer. An exception is the case when all three angular momentum projections are zero, which occurs when $l = 0$.

Does the formula $M^2 = h^2 l \, (l+1)$ cover all the eigenvalues of the angular momentum square for integral values of l? If we consider only the spatial motion of a particle with three degrees of freedom, then there exist only integral l's. This is known as the *orbital angular momentum*. The term was borrowed from Bohr's old theory, which assumed that particles travel along orbits. In the more general case, when a particle also has rotation-type internal degrees of freedom, half-integral l's are possible, but for the time being we shall consider only orbital angular momentum.

It follows from Eq. (24.33) that an application of the operator $(\mathbf{a}_i \cdot \nabla)$ to the expression appearing in it always results in the appearance of scalar products of the form $(\mathbf{a}_i \cdot \mathbf{r})$, or of constant numbers $(\mathbf{a}_i \cdot \mathbf{a}_n)$. They are all single-valued functions of the coordinates. But such homogeneous functions can be obtained only for integral powers of l. Nonintegral powers of l in a homogeneous function signify nonintegral powers of the coordinates, that is, the single-valuedness disappears. But the eigenfunction of the operator $\hat{M}^2$ is the amplitude of the probability that for $M^2 = h^2 l \, (l+1)$ a particle has a polar angle ϑ and azimuth φ:

$$dw = |\psi_l(\vartheta, \varphi)|^2 \, d\Omega$$

This function must, by definition, be single valued. Therefore non-integral values of l for the orbital angular momentum are precluded.

From Eq. (24.26), the operator of the square of the angular momentum commutes with the operator of any of its projections. The same is apparent from (24.28) if we represent the angular momentum square in the form

$$\hat{M}^2 = -\frac{h^2}{\sin\vartheta}\,\frac{\partial}{\partial\vartheta}\sin\vartheta\,\frac{\partial}{\partial\vartheta} + \frac{\hat{p}_\varphi^2}{\sin^2\vartheta} \tag{24.35}$$

with the help of (24.9). Since angle φ is not explicitly involved in (24.35), $\hat{M}^2$ commutes with $\hat{p}_\varphi$.

But this means that the quantity M^2 and the quantity p_φ can exist in the same state, that is, states with definite values of M^2 and p_φ can be described in terms of the same eigenfunction. The eigenfunction of $\hat{p}_\varphi$ is immediately apparent from (24.9). Writing the eigenvalue equation for p_φ, we have

$$\hat{p}_\psi\psi = \frac{h}{i}\,\frac{\partial\psi}{\partial\varphi} = p_\varphi\psi \tag{24.36}$$

whence

$$\psi = e^{i\varphi p_\varphi/h} \tag{24.37}$$

For arbitrary vectors $\mathbf{a}_i$ the eigenfunction ψ does not correspond to (24.37). But it is not difficult to choose these vectors in a way such that an operation with $(\mathbf{a}_i \cdot \nabla)$ would reduce the eigenfunction to the required form. For this we must select only two types of $\mathbf{a}_i$. We take $l - k$ of these vectors as unit vectors along the z axis, and the remaining k vectors in the form of a linear combination of unit vectors along the x and y axes:

$$\mathbf{a}_\pm = \mathbf{n}_x \pm i\mathbf{n}_y$$

Then $(\mathbf{a}_+ \cdot \nabla) = (\partial/\partial x) \pm i\,(\partial/\partial y)$. The operation of $(\mathbf{n}_z \cdot \nabla) = (\partial/\partial z)$ on r^{-1} yields z/r^3. Repeated application of the same operator with respect to r^{-3} again yields z and unity in the numerator, operating on the z carried over from the preceding differentiation. Applying the operator $(\partial/\partial x) + i\,(\partial/\partial y)$ to r^{-1}, we obtain

$$\left(\frac{\partial}{\partial x} + i\,\frac{\partial}{\partial y}\right)\frac{1}{r} = -\frac{x + iy}{r^3} = -\frac{\sin\vartheta\,e^{i\varphi}}{r^2}$$

Obviously, repeated operation with the same operator multiplies the obtained expression once again by $e^{i\varphi}$, besides leading to the appearance of powers of $\sin\vartheta$. Indeed, once again differentiating the preceding expression in the same way, we find that

$$-\left(\frac{\partial}{\partial x} + i\,\frac{\partial}{\partial y}\right)\frac{x + iy}{r^3} = -\frac{1 + i^2}{r^3} + \frac{3(x + iy)^2}{r^5} = \frac{3\sin^2\vartheta\,e^{2i\varphi}}{r^5}$$

and so on.

Thus, if the operator $(\partial/\partial x) + i\,(\partial/\partial y)$ is applied k times, the result is proportional to $e^{ik\varphi}$. This is the eigenfunction of the operator $\hat{p}_\varphi$, if the eigenvalue $p_\varphi = hk$. But since we have found that any eigenvalue of $\hat{M}^2$ corresponds to the eigenvalue $h^2l\,(l + 1)$ for integral l, and differentiation $(\partial/\partial x) + i\,(\partial/\partial y)$ is part of all the differentiations of the form $(\mathbf{a}\cdot\nabla)$, we find that k, too, is of necessity an integer. Obviously, it cannot be greater than l. Furthermore, if the operator $(\partial/\partial x) - i\,(\partial/\partial y)$ is taken, then the eigenfunction contains $-ik\varphi$ in the exponent, though again in absolute value k does not exceed l. Thus, we have obtained all the eigenvalues $p_\varphi = hk$:

$$-hl \leqslant p_\varphi \leqslant hl, \quad -l \leqslant k \leqslant l \qquad (24.38)$$

In the process we developed the simultaneous eigenfunctions of the operator of the square of the angular momentum and its projection, $\hat{M}_z$. In mathematics these are known as *spherical functions*. We have defined them as follows:

$$Y_l^{\pm k} = r^{l+1}\left(\frac{\partial}{\partial z}\right)^{l-k}\left(\frac{\partial}{\partial x} \pm i\,\frac{\partial}{\partial y}\right)^k \frac{1}{r} \qquad (24.39)$$

They satisfy two equations simultaneously:

$$\hat{M}^2 Y_l^{\pm k} = h^2 l\,(l + 1)\,Y_l^{\pm k} \qquad (24.40a)$$

$$\hat{M}_z Y_l^{\pm k} \equiv \hat{p}_\varphi Y_l^{\pm k} = \pm\,hk Y_l^{\pm k} \qquad (24.40b)$$

The Stern-Gerlach Experiment. That the angular momentum projections are integers is confirmed by direct experiment. The idea of the experiment consists in the following: a direct relationship exists between the orbital angular momentum projection and the magnetic moment projection (see (17.30)):

$$\mu_z = \frac{e}{2mc}\,M_z \qquad (24.41)$$

A narrow beam of vapour of a substance under investigation is passed between the poles of an electromagnet in a strongly inhomogeneous field; to achieve this, one of the poles should be made tapered. The particles (in the Stern-Gerlach experiment they are atoms) enter the field parallel to the edge of the taper, that is, they move in a direction perpendicular to the plane of the lines of force of the field. The plane of symmetry of the field passes through the edge of the taper and the direction of motion of the particles. We assume the z axis to be perpendicular to the edge of the taper and to lie in the plane of symmetry of the field. If the angular momentum of the atoms has only discrete, integral projections on the z axis, then the magnetic moment of the atoms has only several definite values

corresponding to the angular momentum. The deflecting force acting on a particle possessing magnetic moment in a magnetic field is, by (17.35)

$$(\mu \cdot \nabla) H_z = \mu_z \frac{\partial H_r}{\partial z} = k \frac{eh}{2mc} \frac{\partial H_z}{\partial z} \qquad (24.42)$$

In the plane of symmetry of the field, **H** is in the *z*-direction and depends only upon *z*.

Since the angular momentum M_z can only have a definite set of values, the deflecting force acting upon the atoms in the beam also
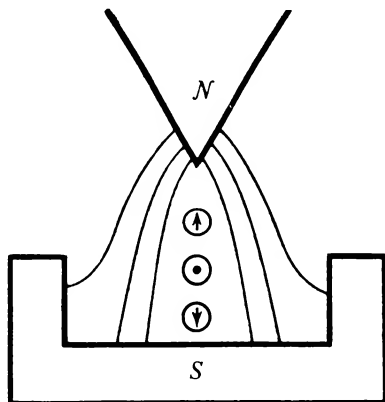
Figure 29

has a very definite value for particles with the respective angular momentum projection M_z. It can be seen from (24.42) that the force is a multiple of $(eh/2mc)(\partial H_z/\partial z)$. Therefore, the particles in the beam experience only those deflections in the magnetic field which correspond to the possible values of the force (24.42). In other words, a beam of particles in a deflecting magnetic field does not take the form of a continuous fan, as could be expected according to the classical theory, but separates into as many discrete beams as the number of values *k* takes.

The number *k*, as we have just shown (see (24.38)), lies between the limits from $-l$ to l, that is, it acquires $2l + 1$ values. Figure 29 presents the separation pattern for the case of $l = 1$. As for the number *l* itself, it does not vary in the beam and corresponds to the state of the atoms in it with the least possible energy. This state is associated with a definite value of the square of angular momentum (in the classical analogy—with the "centrifugal energy", as in Section 5) and usually does not vary as the substance evaporates. The thermal energy of evaporation is insufficient to alter the energy

state of an atom, therefore all the atoms travel with the same value of l but with random distributions of the projection of angular momentum k.

The fact that k is restricted to integers is associated with the fact that two angular momentum projections cannot exist simultaneously.

This is easily linked with the Stern-Gerlach experiment. Indeed, the direction of the magnetic field (or the z axis) is taken to be completely arbitrary. All permitted values of the projection of the angular momentum on this axis are equiprobable. We could measure the angular momentum projections on a certain spatial axis and then pass the same beams through a magnetic field at a very small angle with the field in which the first measurement was carried out. Both measurements will, naturally, yield integers for the projections. But the same vector cannot simultaneously have integral projections on infinitely close, but in other respects random, directions: when the first measurement was undertaken, the angular momentum had projections only on the first direction of the field, and correspondingly, in the second measurement only on the second direction of the field.

We find that just as the coordinate and linear momentum do not exist simultaneously, so two angular momentum projections do not exist in the same state.

EXERCISES

1. Calculate the commutators $[\hat{M}_x \hat{p}_x]$, $[\hat{M}_x \hat{p}_y]$, $[\hat{M}_x \hat{p}_z]$, $[\hat{M}_x \hat{x}]$, $[\hat{M}_x \hat{y}]$, $[\hat{M}_z z]$, $[\hat{M}^2 \hat{p}_x]$, $[\hat{M}^2 \hat{p}^2]$, and $[\hat{M}^2 \hat{r}^2]$, where the brackets symbolise commutators of the operators inside them.

2. Prove that $\hat{p}_x f(\hat{x}) - f(\hat{x}) \hat{p}_x = (h/i)(\partial f/\partial x)$, where $f(\hat{x})$ is a function of $\hat{x}$.

3. Verify that $[\hat{a}[\hat{b}\hat{c}]] + [\hat{b}[\hat{c}\hat{a}]] + [\hat{c}[\hat{a}\hat{b}]] = 0$ (the Jacobi identity).

4. If the commutation of two operators, $[\hat{a}\hat{b}]$, is a number and not an operator, then

$$e^{\hat{a}+\hat{b}} = e^{[\hat{a}\hat{b}]} e^{\hat{a}} e^{\hat{b}}$$

where $e^{\hat{a}} = 1 + \frac{\hat{a}}{1!} + \frac{\hat{a}^2}{2!} + \frac{\hat{a}^3}{3!} + \ldots$. Prove this statement.

5. Prove that in going over to spherical coordinates the following equations hold true:

$$\hat{M}_\pm \equiv (\hat{M}_x \pm i\hat{M}_y) = h e^{i\varphi} \left(\frac{\partial}{\partial\vartheta} \pm i \cot \vartheta \frac{\partial}{\partial\varphi} \right)$$

25

EXPANSIONS IN WAVE FUNCTIONS

The Superposition Principle. One of the most fundamental ideas of quantum mechanics is that its equations are linear with respect to the wave function ψ. This result proceeds from the whole set of facts that confirm the correctness of quantum mechanics, in the same way as an analogous result in classical electrodynamics (see Sec. 15), which is also a generalization of experience.

For example, the diffraction of electrons shows that the amplitudes of wave functions are combined in the same simple way as the amplitudes of waves in optics; diffraction maxima and minima are situated at the same positions, determined only by the phase relationships, independently of the wave intensities. All this points to the linearity of wave equations; the solutions of nonlinear equations behave in an entirely different manner.

The sum of two solutions of a linear equation again satisfies the same equation. It follows from this that any solution of a wave equation can be represented in the form of a certain set of standard solutions, similar to the way that, in Section 19, a travelling nonperiodic wave was represented by a set of travelling harmonic waves.

The statement concerning the possibility of representing a single wave function in terms of the sum of other wave functions is called the *superposition principle.*

Hermitian Operators. Wave functions can usually be represented with the aid of the sum of other wave functions which are eigenfunctions of certain quantum mechanical operators. In the present section it will be shown how such expansions are performed. First of all, however, it is necessary to establish certain general properties of operators whose eigenvalues are measurable physical quantities.

We have already had an example of measuring a physical quantity in the Stern-Gerlach experiment. The total number of particle beams resulting from the splitting of the initial beam in a magnetic field defines the number l, that is, the angular momentum square, while the number of a given individual beam in the assembly of $2l + 1$ beams defines the projection of the angular momentum $\hbar k$ on the selected axis.

Obviously, the measured eigenvalues must be real numbers, although the operators themselves may depend explicitly upon $i = \sqrt{-1}$ (see (24.3) and (24.9)). We shall consider the equations for the eigenfunctions of the operator $\hat{\lambda}$ and another equation involv-

ing its complex conjugate:

$$\hat{\lambda}\psi = \lambda\psi \tag{25.1a}$$

$$\hat{\lambda}^*\psi^* = \lambda^*\psi^* \tag{25.1b}$$

We must find the condition for which the eigenvalues are real numbers, $\lambda^* = \lambda$.

To do this, we multiply (25.1a) by ψ^* and (25.1b) by ψ, integrate over the whole range of x (upon which the operators may depend), and subtract one from the other. This yields

$$\int (\psi^*\hat{\lambda}\psi - \psi\hat{\lambda}^*\psi^*)\, dx = (\lambda - \lambda^*) \int \psi^*\psi\, dx$$

But the integral of $\psi^*\psi = |\psi|^2$ cannot be equal to zero, since $|\psi|^2$ is an essentially positive quantity.

The eigenvalue of the observed quantity, λ, is by definition real, that is $\lambda = \lambda^*$; therefore we arrive at the relation

$$\int (\psi^*\hat{\lambda}\psi - \psi\hat{\lambda}^*\psi^*)\, dx = 0 \tag{25.2}$$

Equation (25.2) must hold not only for the pair of complex conjugate eigenfunctions of the operator $\hat{\lambda}$, that is $\psi(\lambda,\ x)$ and $\psi^*(\lambda,\ x)$, but for any pair of functions $\chi^*(x)$, $\psi(x)$ which satisfy the same conditions (finiteness, uniqueness, continuity) as the eigenfunctions $\psi(\lambda,\ x)$:

$$\int (\chi^*\hat{\lambda}\psi - \psi\hat{\lambda}^*\chi^*)\, dx = 0 \tag{25.3}$$

The necessity of such a more comprehensive formula in comparison with (25.2) will be explained later in this section. An operator for which Eq. (25.3) is satisfied is called a *Hermitian operator* and the corresponding property of such operators is termed *Hermiticity*.

We shall now verify whether the operators introduced up till now are Hermitian. Take, for example, $M_z = (h/i)\,(\partial/\partial\varphi)$, for which we obtain

$$\int\limits_0^{2\pi} \chi^*\hat{M}_z\psi\, d\varphi = \int\limits_0^{2\pi} \chi^*\,\frac{h}{i}\,\frac{\partial\psi}{\partial\varphi}\, d\varphi = \frac{h}{i}\,\chi^*\varphi \Big|_0^{2\pi} - \int\limits_0^{2\pi} \psi\,\frac{h}{i}\,\frac{\partial\chi^*}{\partial\varphi}\, d\varphi$$

The eigenfunctions of the operator $\hat{M}_z$ were found in the preceding section. They are equal to $e^{ik\varphi}$. For integral k, such functions satisfy the single-valuedness condition $\psi(\varphi) = \psi(\varphi + 2\pi)$, so that the functions χ^* and ψ substituted into the integral must also be single valued. The integrated part of the expression therefore vanishes when the limits are substituted, whence follows the Hermiticity

of $\hat{M}_z$:

$$\int\limits_0^{2\pi} \chi^* \hat{M}_z \psi \, d\varphi = \int\limits_0^{2\pi} \psi \hat{M}_z^* \chi^* \, d\varphi$$

in agreement with the general requirement (25.3).

In the case of $\hat{M}^2$ we must take, instead of $d\varphi$, the solid angle element $d\Omega$. Then the Hermiticity of $\hat{M}^2$ is proved by double integration by parts: over φ and over ϑ. For $\hat{\mathscr{H}}$ (the Hamiltonian), dx corresponds to dV. To prove the Hermiticity of $\hat{\mathscr{H}}$ the integration must be performed with the help of the Gauss theorem.

The Orthogonality of Eigenfunctions. An important property of eigenfunctions follows from the Hermiticity of operators. Let us consider the equations for two eigenvalues of the same operator $\hat{\lambda}$:

$$\hat{\lambda}\psi(\lambda, x) = \lambda\psi(\lambda, x) \tag{25.4a}$$

$$\hat{\lambda}^*\psi^*(\lambda', x) = \lambda'\psi^*(\lambda', x) \tag{25.4b}$$

We multiply (25.4a) by $\psi^*(\lambda', x)$, and (25.4b) by $\psi(\lambda, x)$, integrate over x, and subtract one from the other to get

$$\int [\psi^*(\lambda', x)\,\hat{\lambda}\psi(\lambda, x) - \psi(\lambda, x)\,\hat{\lambda}^*\psi^*(\lambda', x)]\,dx$$

$$= (\lambda - \lambda') \int \psi^*(\lambda', x)\,\psi(\lambda, x)\,dx \tag{25.5}$$

The left-hand side of this equation vanishes in accordance with the general requirements for Hermiticity (25.3). Therefore, if $\lambda' \neq \lambda$, the following integral must vanish:

$$\int \psi^*(\lambda', x)\,\psi(\lambda, x)\,dx = 0 \tag{25.6}$$

This property of wave functions is called *orthogonality*.

Sometimes several quantities λ, ν, etc. may correspond to the same state of a system. For that the operators $\hat{\lambda}$, $\hat{\nu}$, ... must commutate. For example, for free motion of a particle there exist p_x, p_y, and p_z or, as we have shown before, $\hat{M}^2$ can have an eigenvalue together with $\hat{M}_z$. Then we develop functions that are simultaneous eigenfunctions with respect to all operators, the eigenfunctions being of the type $e^{i(p_x x + p_y y + p_z z)/\hbar}$ or Y_l^k. In the most general case the eigenvalue equations

$$\hat{\lambda}\psi(\lambda, \nu; x) = \lambda\psi(\lambda, \nu; x)$$

$$\hat{\nu}\psi(\lambda, \nu; x) = \nu\psi(\lambda, \nu; x)$$

must hold if

$$[\hat{\lambda}\hat{\nu}] \equiv \hat{\lambda}\hat{\nu} - \hat{\nu}\hat{\lambda} = 0$$

For such functions

$$\int \psi^* (\lambda', \nu'; x)\, \psi (\lambda, \nu; x)\, dx = 0 \qquad (25.7)$$

if $\lambda' \neq \lambda$ or $\nu' \neq \nu$.

Expansion in Eigenfunctions. Let us suppose that the eigenfunctions of a certain operator $\hat{\lambda}$ are known. These functions satisfy, in addition to the equation $\hat{\lambda}\psi = \lambda\psi$, certain requirements associated with the conditions of the eigenfunction problem: they are finite, continuous, single valued, and so forth. Then, in accordance with the superposition principle, any function $\psi (x)$ which satisfies the same requirements may be represented as the sum of the eigenfunctions of the operator $\hat{\lambda}$:

$$\psi (x) = \sum_{\lambda'} c_{\lambda'} \psi (\lambda', x) \qquad (25.8)$$

We shall show how to determine the expansion coefficients $c_{\lambda'}$. For this we multiply both sides of the equation by $\psi^* (\lambda, x)$ and integrate over x:

$$\int \psi^* (\lambda, x)\, \psi (x)\, dx = \sum_{\lambda'} c_{\lambda'} \int \psi^* (\lambda, x)\, \psi (\lambda', x)\, dx \qquad (25.9)$$

In accordance with the orthogonality condition all the integrals in the right-hand side of (25.9) vanish except the one with $\lambda' = \lambda$. Consequently, there remains the equation

$$\int \psi^* (\lambda, x)\, \psi (x)\, dx = c_{\lambda} \int \psi^* (\lambda, x)\, \psi (\lambda, x)\, dx$$

$$= c_{\lambda} \int |\psi (\lambda, x)|^2\, dx$$

We shall consider the eigenfunctions $\psi (\lambda, x)$ normalized to unity, that is $\int |\psi|^2\, dx = 1$ (see (23.17)). The normalization condition can be written together with the orthogonality condition (25.6) in the form of one equation with the help of the symbol $\delta_{\lambda\lambda'}$, where $\delta_{\lambda\lambda'} = 0$ for $\lambda \neq \lambda'$ and $\delta_{\lambda\lambda'} = 1$ for $\lambda = \lambda'$:

$$\int \psi^* (\lambda', x)\, \psi (\lambda, x)\, dx = \delta_{\lambda\lambda'} \qquad (25.10a)$$

where $\delta_{\lambda\lambda'}$ is, of course, not a tensor but simply a symbol with the stated properties.

For the expansion coefficient c_λ we obtain

$$c_\lambda = \int \psi^* (\lambda, \ x) \, \psi \, (x) \, dx \qquad (25.11a)$$

In the case when we have a system of commutative operators $\hat{\lambda}, \hat{\nu}$, Eq. (25.11a) is directly generalized to

$$c_{\lambda, \, \nu} = \int \psi^* (\lambda, \ \nu; \ x) \, \psi \, (x) \, dx \qquad (25.11b)$$

if

$$\int \psi^* (\lambda', \ \nu'; \ x) \, \psi \, (\lambda, \ \nu; \ x) \, dx = \delta_{\lambda\lambda'} \delta_{\nu\nu'} . \qquad (25.10b)$$

Thus, a state[1] $\psi \, (x)$ in which the quantity λ has no definite value because $\psi \, (x)$ is not an eigenvalue of $\hat{\lambda}$, is represented as the sum of states with strictly defined eigenvalues λ. The component of the wave function corresponding to a certain value of λ is

$$c_\lambda \psi \, (\lambda, \ x) \qquad (25.12)$$

It represents the probability amplitude of the given value of the quantity λ in the state $\psi \, (x)$. To find the probability w_λ of the occurrence of the quantity λ, we must eliminate the dependence on x, since x and λ do not exist in the same state.

For this we determine the probability density of the state with the given λ, that is $| \, c_\lambda \, |^2 \, | \, \psi \, (\lambda, \ x) \, |^2$, and integrate over x. From the normalization condition for eigenfunctions we obtain

$$w_\lambda = | \, c_\lambda \, |^2 \int | \, \psi \, (\lambda, \ x) \, |^2 \, dx = | \, c_\lambda \, |^2 \qquad (25.13)$$

Let us now verify that the quantities $| \, c_\lambda \, |^2$ possess the basic property of probability: their sum over λ is equal to unity if the function $\psi \, (x)$ itself satisfies the normalization condition (23.17). Indeed, making use of the orthogonality condition (25.10a), we obtain

$$1 = \int | \, \psi \, (x) \, |^2 \, dx = \int \left(\sum_\lambda c_\lambda^* \psi^* (\lambda, \ x) \sum_{\lambda'} c_{\lambda'} \psi \, (\lambda', \ x) \right) dx$$

$$= \sum_{\lambda\lambda'} c_\lambda^* c_{\lambda'} \delta_{\lambda\lambda'} = \sum_\lambda | \, c_\lambda \, |^2$$

Comparing with (25.13), we have

$$\sum_\lambda w_\lambda = \sum_\lambda | \, c_\lambda \, |^2 = 1 \qquad (25.14)$$

---

[1] For the sake of brevity it is customary to speak of a state $\psi \, (x)$ instead of "a state with wave function $\psi \, (x)$".

Thus, the coefficient c_λ should be regarded as a probability amplitude similar to $\psi(x)$. But $\psi(x)$ is connected with the probability of finding a particle with coordinate x independently of λ, while c_λ with the probability of finding it with the given value of λ independently of x.

Now we can return to the meaning of the condition for the Hermiticity of operators, (25.3). From this condition derives the orthogonality of eigenfunctions with different values of λ. If a system is in a state with a certain eigenvalue λ, the probability amplitude of another value $\lambda' \neq \lambda$ occurring in that state vanishes, because it follows from (25.11a) that

$$c_{\lambda'} = \int \psi^*(\lambda', x)\, \psi(\lambda, x)\, dx = 0 \tag{25.15}$$

provided $\lambda' \neq \lambda$. Thus, from the Hermiticity of an operator follows the reality of its eigenvalues and the possibility of "pure states" with exactly defined values of the corresponding quantity.

An expansion of a function $\psi(x)$ in the eigenvalues of operator $\hat{\lambda}$ is very like the expansion of a vector in unit vectors directed along the axes of a Cartesian coordinate system. The part of these unit vectors is played by the eigenfunctions of the operator, $\psi(\lambda, x)$, and the part of the vector projections on the axes by the expansion coefficients, c_λ. The normalization condition is similar to the choice of unit vectors for the expansion $(n^{(i)})^2 = 1$, and if the expanded vector is of unit length, that is, itself normalized, the sum of its projections is also unity, like the condition $\sum |c_\lambda|^2 = 1$.

The definition of the magnitude of the projection of a vector A on an axis in tensor notation is

$$A_i = A_\alpha n_\alpha^{(i)}$$

Then

$$\mathbf{A} = \sum_i A_i \mathbf{n}^{(i)}$$

These two formulas should be compared with (25.8) and (25.11a). Summation over the tensor index α is similar to integration over x; summation over the unit vector's number i is similar to summation over λ. We as it were have a vector $\psi(x)$ in a space with an infinite number of dimensions instead of the three dimensions of Euclidean space. Besides, the "length" of such a vector is defined not as the sum of the squares of its components but as the sum of the squares of their moduli, since the vector is complex.

The comparison of $\psi(x)$ with a vector can be continued. For that we refer to Eq. (9.9a), which is used to find the principal axes and principal values of the inertia tensor $I_{\alpha\beta}$. This equation is similar to (24.12), from which we define the eigenvalues of the operator $\hat{\lambda}$

and the eigenfunctions of $\psi(\lambda, x)$, which, as was just pointed out, are analogous to unit vectors along the Cartesian axes. The orthogonality of wave functions has an analogy in the perpendicularity of the principal axes of the inertia tensor:

$$n_\alpha^{(i)} n_\alpha^{(k)} = \delta_{ik}$$

It was shown in Section 9 that the perpendicularity of the principal axes of the inertia tensor derives from the symmetry of the tensor, $I_{\alpha\beta} = I_{\beta\alpha}$. In a space of complex vectors the requirement in place of symmetry is the Hermiticity of the operators. We shall later show that symmetry and Hermiticity are written in very similar fashion.

The mathematical concept of a complex space similar to Euclidean space was of great help in formulating the laws of quantum mechanics. It is known in mathematics as Hilbert space.

Expansion in Eigenfunctions of the Angular Momentum Projection. We shall explain the meaning of expansions in eigenfunctions with the example of the Stern-Gerlach experiment. A beam of atoms splits into a certain number of separate beams according to the number of projections of the angular momentum on the magnetic field, $M_z = hk$. If the greatest value of the projection is equal to hl, then k, as pointed out before, assumes $2l + 1$ values, from $-l$ to l, changing by unity.

The eigenfunction corresponding to $M_z = hk$ is

$$\psi(k, \varphi) = \frac{1}{(2\pi)^{1/2}} e^{ik\varphi} \tag{25.16a}$$

where the factor $(2\pi)^{-1/2}$ is introduced for normalization:

$$\int_0^{2\pi} |\psi|^2 \, d\varphi = 1$$

If each of the separate beams is once again passed through a magnetic field parallel to the z axis, there is no further splitting; this is because M_z in these beams has a single definite value and not the whole set of values in the range $-l \leqslant k \leqslant l$, as was the case in the initial beam. From this the meaning of the orthogonality of eigenfunctions is very well seen. If a particle is found in a beam corresponding to a given value of k, then the probability of finding it in a beam with a different value of the projection $M_z = hk' \neq hk$ is equal to zero.

From the general rule, the probability equals the square of the modulus of the expansion coefficient $c_{k'}$ of the function $\psi(k, \varphi)$

in functions of $\psi(k', \varphi)$, that is, according to the general formula

$$c_{k'} = \int_0^{2\pi} \psi^*(k, \varphi)\, \psi(k', \varphi)\, d\varphi = \frac{1}{2\pi} \int_0^{2\pi} e^{i\varphi(h'-k)}\, d\varphi$$

$$= \frac{1}{2\pi} \frac{e^{i\varphi(k'-k)}}{i(k'-k)}\Big|_0^{2\pi} = \begin{cases} 0, & k' \neq k \\ 1, & k' = k \end{cases}$$

If the second magnetic field is along the x axis, then splitting will again occur due to the component of angular momentum M_x, which does not exist simultaneously with M_z. The number of splitting components is again equal to $2l + 1$, since it is determined by the maximum angular momentum projection l. This quantity cannot depend upon the direction of the magnetic field, and is related only to the atomic states in the original beam.

The eigenfunctions of $\hat{M}_x$ are

$$\psi(k_1, \omega) = \frac{1}{(2\pi)^{1/2}} e^{ik_1\omega} \tag{25.16b}$$

where $-l \leqslant k_1 \leqslant l$, and ω is the angle of rotation about the x axis.

Functions (25.16a) and (25.16b) do not coincide, which is a natural consequence of their being functions of noncommuting operators.

As a result of magnetic splitting in a field directed along the x axis, a beam with given value of k is split into $2l + 1$ beams with definite values of M_x. Hence, the function (25.16a) is represented as the sum of functions (25.16b):

$$\psi(k, \varphi) = \sum_{k_1=-l}^{l} c_{k_1} \psi(k_1, \omega) \tag{25.17}$$

The square of the modulus, $|c_{k'}|^2$, indicates the proportion of particles from a beam with a given value of k that will occur in the beam corresponding to the angular momentum projection on the x axis equal to hk_1.

If the second magnetic field is not oriented along the x axis and makes a small angle with the initial axis z, then the largest square of the modulus among expansion coefficients in (25.17) occurs for the one for which the value of the projection on the new axis (k_1) closely approximates the projection on the old axis (k). The other coefficients are small.

The Wave Function and the Measurement of Quantities. The foregoing example of determining the angular momentum projections explains the role of measurement processes in quantum mechanics. Before the Stern-Gerlach experiment was carried out with the initial beam, that is, before the particles passed through a magnetic field,

nothing at all could be said of the value of their angular momentum projections. Depending on the orientation of the field the angular momentum projections were obtained either on the z axis or on the x axis, but, of course, not both simultaneously.

From this can be seen the special part that measuring instruments play in quantum mechanics, which is substantially different from their part in classical (nonquantum) physics. For whereas classical measurements have an infinitesimal effect on the object being measured, measurements carried out on microscopic entities may affect them so greatly as to simply preclude other simultaneous measurements.

For example, when measuring the angular momentum projection on the z axis, it is impossible simultaneously to measure another projection. This follows convincingly from the Stern-Gerlach experiment. In measuring the coordinate of a particle, we cannot at the same time measure its linear momentum in the corresponding direction. To measure the coordinate to an accuracy of Δx, the particle must be passed through a slit of width Δx, but then diffraction produces an inaccuracy in the linear momentum of $\Delta p_x \geqslant$ $\geqslant 2\pi h/\Delta x$. Here the measuring instrument is a slitted screen. Thus, in considering measuring processes we must reckon with the measuring instrument, which is a classical entity, like a magnet or a slitted screen.

The determination of any physical quantity is inseparably linked with the method of measuring it. In classical physics the connection is less apparent, since the measurement has an infinitesimal effect on the measured object. In quantum mechanics the reverse is true: as a rule the state of the measured object after the measurement differs from what it was prior to the measurement. In accordance with the uncertainty principle, it is therefore meaningless to carry out simultaneous measurements of certain quantities. Thus, the principle is substantiated in analysing the measuring operation. In this sense, the action quantum h as it were measures the effect of the instrument on the micro-object; for example, it states the uncertainty of the linear momentum appearing in the particle's passage through the slot.

It should not be imagined, however, that the experimenter "interferes" in some way with the results of physical measurements. As a result of experiments on a large number of identical objects we can determine the state they were in prior to the measurement quite independently of the method of measurement. All the experimenter does is select the method (for example, the direction of the magnetic field in the beam-splitting experiment); but he then obtains a very definite number of splitting components and a definite intensity of every beam. If a single passage of a beam through a magnetic field is being studied, the experimenter always has $2l + 1$ splitting

components, from which he draws the conclusion that in the given beam the angular momentum projection had no definite value.

Now let one of the beams, with a given value of k, be passed into another chamber, and let a second experimenter perform the Stern-Gerlach experiment again.

If the magnetic field in the second experiment is oriented arbitrarily, he will obtain, as mentioned, $2l + 1$ beams, but by varying the direction of the magnetic field he can achieve a state in which no further splitting occurs. This happens when the second magnetic field is directed parallel to the initial one. Hence, the second experimenter will conclude that the particles in the incoming beam have a wave function of the form (25.16a). Here the measurement makes it possible to establish that prior to the measurement process the particles were in some definite "pure" state.

If the field is so oriented that a new splitting of the particle beam is observed in the second experiment, then the wave function in all of the resultant beams will not coincide with the initial function. But there will be a clearcut relationship (25.17) between them, which is quite independent of the will of the experimenter making the measurements.

Quantum mechanics does not make the result of a measurement process subjectivistic; it simply restricts the possibility of simultaneously carrying out certain measurements. Either one or another is possible, but not both at once. This statement is known as the *complementarity principle*. In effect it is equivalent to the uncertainty principle

Mean Values in Quantum Mechanics. We have seen that in quantum mechanics measurement of a quantity need not necessarily yield a single strictly defined value. The probability of a certain value being obtained in a measuring process is unity, that is certainty, only in the "pure" state. In the most general case, the value of a measured quantity λ is obtained with the probability w_λ. Let us find the mean value of the measured quantity according to the definition

$$\langle \lambda \rangle = \sum_\lambda \lambda w_\lambda \qquad (25.18)$$

Substituting w_λ from (25.13) and c_λ from (25.11a), we have

$$\langle \lambda \rangle = \sum_\lambda \lambda \, | \, c_\lambda \, |^2 = \sum_\lambda \lambda c_\lambda c_\lambda^* = \sum_\lambda \lambda c_\lambda^* \int \psi(x) \, \psi^*(\lambda, \, x) \, dx$$

Making use of the condition that the eigenvalues are real, we replace the product $\lambda \psi^*(\lambda, \, x)$ under the integral by $\hat{\lambda}^* \psi^*(\lambda, \, x)$, and first

sum and then integrate. Then we obtain

$$\langle \lambda \rangle = \int \psi(x) \sum_\lambda c_\lambda^* \hat{\lambda}^* \psi^*(\lambda, \ x) \, dx$$

But the operator $\hat{\lambda}^*$ does not depend upon any definite value of λ (for example, if $\lambda = p_x$, then $\hat{\lambda}^* = -(h/i)(\partial/\partial x)$). Therefore we take $\hat{\lambda}^*$ outside the sign of the summation to get

$$\langle \lambda \rangle = \int \psi(x) \, \hat{\lambda}^* \sum_\lambda c_\lambda^* \psi^*(\lambda, \ x) \, dx$$

But the sum $\sum c_\lambda^* \psi^*(\lambda, \ x) = \psi^*(x)$, since this is an equation which is a complex conjugate of (25.8). Therefore

$$\langle \lambda \rangle = \int \psi(x) \, \hat{\lambda}^* \psi^*(x) \, dx$$

Finally, using the Hermiticity of $\hat{\lambda}$, that is, Eq. (25.3), we obtain the required expression for the mean value of λ:

$$\langle \lambda \rangle = \int \psi^*(x) \, \hat{\lambda} \psi(x) \, dx \tag{25.19}$$

Thus, in order to calculate the mean value of λ in a state $\psi(x)$, it is not necessary to know the eigenvalues of $\hat{\lambda}$, since it is sufficient to calculate the integral (25.19).

If state $\psi(x)$ is not a "pure" eigenstate of $\hat{\lambda}$, each measurement puts the particle in another state. However, given a sufficient number of particles in the same initial state, if we consecutively carry out the measurements on each one, we can obtain the value of $\langle \lambda \rangle$ in the initial state up to any accuracy. This value can always be reproduced, provided, of course, that the measurements are not carried out on particles already once measured, but on a "fresh" batch of particles in the same state as that which was obtained prior to the first series of measurements.

Proof of the Uncertainty Relations for $\langle \Delta x \rangle \, \langle \Delta p \rangle$. We define

$$\langle \Delta x \rangle^2 \equiv \langle (x - x_0)^2 \rangle \tag{25.20a}$$

and

$$\langle \Delta p_x \rangle^2 \equiv \langle (p_x - p_{x0})^2 \rangle \tag{25.20b}$$

To simplify the notation we put $x_0 = 0$ and $p_{x0} = 0$. Let us consider the following integral of an essentially positive quantity:

$$I = \int |(a\hat{x} + ib\hat{p}_x) \psi|^2 \, dx \geqslant 0$$

where a and b are real numbers. Expanding the integrand, we obtain

$$I = a^2 \int \psi^* \hat{x}^2 \psi \, dx + b^2 \int (\hat{p}_x^* \psi^*)(\hat{p}_x \psi) \, dx$$

$$- iab \int (\hat{x}\psi)(\hat{p}_x^* \psi^*) \, dx + iab \int (\hat{p}_x \psi)(\hat{x}\psi^*) \, dx$$

We take advantage of the Hermiticity of $\hat{p}_x$ and transform the terms involving $\hat{p}_x^* \psi^*$ in such a way that there appears the commutator $\hat{p}_x \hat{x} - \hat{x}\hat{p}_x$, which from (24.17) is equal to h/i. The integral must remain positive at all values of a and b, which is possible only if the factors of a^2, b^2, and ab satisfy the inequality

$$4 \left(\int \psi^* \hat{x}^2 \psi \, dx \right) \left(\int \psi^* \hat{p}_x^2 \psi \, dx \right) \geqslant h^2 \int \psi^* \psi \, dx = h^2$$

After extracting the square root, we obtain

$$\langle \Delta x \rangle \langle \Delta p_x \rangle \geqslant \frac{h}{2} \tag{25.21}$$

which is the lowest estimate of the inaccuracies $\langle \Delta x \rangle$ and $\langle \Delta p_x \rangle$.

It is apparent from the reasoning that (25.21) is valid for any pair of Hermitian operators, provided their commutator is a number.

EXERCISES

1. The mean values of a certain quantity belonging to a state which is at the same time the eigenstate of the operator corresponding to it, coincide with the eigenvalues of that quantity. Show that the eigenvalues of the square of angular momentum are equal to $h^2 l \, (l + 1)$.

Solution. We construct the square of angular momentum:

$$M^2 = M_x^2 + M_y^2 + M_z^2$$

and find the mean value of both sides of the equation:

$$\langle M^2 \rangle = \langle M_x^2 \rangle + \langle M_y^2 \rangle + \langle M_z^2 \rangle$$

Next we make use of the fact that the mean value of the square of any angular momentum projection is the same:

$$\langle M_x^2 \rangle = \langle M_y^2 \rangle = \langle M_z^2 \rangle, \qquad \langle M^2 \rangle = \langle 3M_z^2 \rangle$$

The mean value of the square of the angular momentum projection can be determined from the fact that all its projections are equiprobable:

$$\langle M_z^2 \rangle = \frac{1}{2l+1} \sum_{-l}^{l} h^2 k^2 = \frac{h^2}{2l+1} \frac{l\,(l+1)\,(2l+1)}{3}$$

But since all the considered states with various projections M_z are eigenstates with respect to M^2, we obtain

$$M^2 = \langle M^2 \rangle = h^2 l (l + 1)$$

2. Determine the energy eigenvalues of a quantum symmetric top. (An example of such a top is an ammonia molecule, which has the shape of a pyramid whose base is a regular trigon.)

Solution. The energy of a symmetric top expressed in terms of the angular momentum projections is (see (9.17)):

$$\hat{\mathscr{H}} = \frac{1}{2I_1} (\hat{M}_1^2 + \hat{M}_2^2) + \frac{1}{2I_3} \hat{M}_3^2$$

Passing to $\hat{M}^2$, we have

$$\hat{\mathscr{H}} = \frac{1}{2I_1} (\hat{M}^2 - \hat{M}_3^2) + \frac{1}{2I_3} \hat{M}_3^2$$

Substituting the eigenvalues of the angular momentum and its projections, we finally obtain

$$E = \frac{h^2}{2I_1} [l (l + 1) - k^2] + \frac{h^2 k^2}{2I_3} = \frac{h^2}{2I_1} l (l + 1) + \frac{h^2 k^2}{2} \left(\frac{1}{I_3} - \frac{1}{I_1} \right)$$

26

TRANSFORMATION OF INDEPENDENT VARIABLES

In classical mechanics we saw that its laws can be formulated in such a way as to involve the coordinates and linear momenta in equations symmetrically (Sec. 10). In quantum mechanics both quantities do not exist simultaneously. It is, however, legitimate to ask how to effect the transformation from coordinates to momenta, assuming the latter to be independent variables, or, in other words, how to introduce the momenta into the equations in place of the coordinates. We may take some other system of variables in place of the momenta. The only restriction is that all these variables exist in one and the same state.

Matrix Representation of Operators. Suppose we have an equation for the eigenfunctions of an operator $\hat{\lambda}$ written in terms of variables x, in which we must pass to the independent variables v. For the initial equations we take

$$\hat{\lambda}\psi (\lambda, x) = \lambda\psi (\lambda, x) \tag{26.1}$$

$$\hat{v}\psi (v, x) = v\psi (v, x) \tag{26.2}$$

In form they resemble (24.12) and (24.13), but now the operators $\hat{\lambda}$ and $\hat{\nu}$ do not commute.

We expand an eigenfunction of $\hat{\lambda}$ in a series of eigenfunctions of $\hat{\nu}$:

$$\psi(\lambda, x) = \sum_{\nu} c_{\nu}\psi(\nu, x) \qquad (26.3)$$

where the expansion coefficients c_{ν} are given by the general formula (25.11a):

$$c_{\nu} = \int \psi^*(\nu, x)\, \psi(\lambda, x)\, dx \qquad (26.4)$$

Substituting this expansion into Eq. (26.1), we premultiply both sides of the equation by $\psi^*(\nu', x)$ and integrate over x. In the right-hand side, by virtue of (25.10a), we have

$$\int \psi^*(\nu', x) \sum_{\nu} c_{\nu}\psi(\nu, x)\, dx = \sum_{\nu} c_{\nu} \int \psi^*(\nu', x)\psi(\nu, x)\, dx$$

$$= \sum_{\nu} c_{\nu}\delta_{\nu\nu'} = c_{\nu'}$$

and in the left-hand side we introduce the notation

$$\lambda_{\nu'\nu} \equiv \int \psi^*(\nu', x)\, \hat{\lambda}\psi(\nu, x)\, dx \qquad (26.5)$$

In addition, we change the summation index from ν to ν'. Then Eq. (26.1) reduces to the form

$$\sum_{\nu'} \lambda_{\nu\nu'}c_{\nu'} = \lambda c_{\nu} \qquad (26.6)$$

In this form it resembles Eq. (9.9) for finding the principal values of the inertia tensor. Instead of the quantity with two indices, $I_{\alpha\beta}$, we have here another quantity, or rather a set of values, $\lambda_{\nu\nu'}$. This set is more conveniently written as an array

$$\lambda_{\nu\nu'} = \begin{pmatrix} \lambda_{11} & \lambda_{12} & \dots & \lambda_{1\nu'} & \dots \\ \lambda_{21} & \lambda_{22} & \dots & \lambda_{2\nu'} & \dots \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \lambda_{\nu 1} & \lambda_{\nu 2} & \dots & \lambda_{\nu\nu'} & \dots \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \end{pmatrix} \qquad (26.7)$$

which is called a *matrix*. The individual elements of $\lambda_{\nu\nu'}$ are called *matrix elements*.

Thus, in most general form the operator $\hat{\lambda}$ is represented in terms of the arbitrary variables ν as a matrix (26.7). It will be shown further on that the initial form of the operation of the operator $\hat{\lambda}$, (26.1), can also be reduced to the form (26.6).

If (26.6) serves for finding the same eigenvalues λ as (26.1), it is natural to understand the coefficients c_ν appearing in (26.6) as the eigenvalues of $\hat{\lambda}$ in terms of the variables ν, which can be written down as follows:

$$c_\nu \equiv \psi(\lambda, \nu) \tag{26.8}$$

We have thus effected as it were a transition to a new system of coordinate axes in Hilbert space. Equation (26.1) defined a system of eigenvectors of $\hat{\lambda}(x)$, that is $\psi(\lambda, x)$, whereas (26.6) yields a system of eigenvectors $c_\nu = \psi(\lambda, \nu)$ of $\hat{\lambda}(\nu) \equiv \lambda_{\nu\nu'}$. Here $\lambda_{\nu\nu'}$ must be seen as the whole array of matrix elements, the matrix (26.7).

The condition for the Hermiticity of $\lambda_{\nu\nu'}$ can be rewritten in matrix notation very simply. Expressing this condition in the form (25.3), and substituting $\chi = \psi(\nu, x)$ and $\psi = \psi(\nu', x)$ into it, we obtain

$$\int \psi^*(\nu, x) \hat{\lambda} \psi(\nu', x)\, dx = \int \psi(\nu', x) \hat{\lambda}^* \psi^*(\nu, x)\, dx$$

$$= \left(\int \psi^*(\nu', x) \hat{\lambda} \psi(\nu, x)\, dx \right)^*$$

or, using the general notation (26.5) for a matrix element,

$$\lambda_{\nu\nu'} = \lambda^*_{\nu'\nu} \tag{26.9}$$

In this form the condition for the Hermiticity of an operator closely resembles the symmetry condition for a tensor. But since in Hilbert space vectors are complex, a complex conjugation must be performed together with the commutation of indices.

Let us now show how to write in matrix form the result of a successive application of two operators $\hat{\lambda}$ and $\hat{\eta}$ to some vector c_ν. The operation of the first operator is written in the form

$$a_{\nu'} = \sum_\nu \lambda_{\nu'\nu} c_\nu \tag{26.10a}$$

where $a_{\nu'}$ is a new vector (cf. in Section 9: $M_\alpha = I_{\alpha\beta}\omega_\beta$, where ω_β is the angular velocity vector and M_α is the angular momentum vector). Then the application of the next operator $\hat{\eta}$ to $a_{\nu'}$ should be represented in matrix form as

$$b_\mu = \sum_{\nu'} \eta_{\mu\nu'} a_{\nu'} = \sum_{\nu'} \sum_\nu \eta_{\mu\nu'} \lambda_{\nu'\nu} c_\nu = \sum_\nu \left(\sum_{\nu'} \eta_{\mu\nu'} \lambda_{\nu'\nu} \right) c_\nu \tag{26.10b}$$

We see from this that the result of the successive application of two operators to the vector c_ν is the matrix

$$(\eta\lambda)_{\mu\nu} = \sum_{\nu'} \eta_{\mu\nu'} \lambda_{\nu'\nu}, \qquad b_\mu = \sum_\nu (\eta\lambda)_{\mu\nu} c_\nu \tag{26.11}$$

The obtained formula expresses the rule of matrix multiplication. It corresponds to the successive operation of two operators, $\hat{\eta}$ and $\hat{\lambda}$. From formula (26.11) it can be seen that, like the application of operators, multiplication of matrices is noncommutative:

$$(\lambda\eta)_{\mu\nu} = \sum_{\nu'} \lambda_{\mu\nu'}\eta_{\nu'\nu} \neq \sum_{\nu'} \eta_{\mu\nu'}\lambda_{\nu'\nu} = (\eta\lambda)_{\mu\nu} \qquad (26.12)$$

The multiplication of a vector by a number can always be represented in the form

$$ac_\nu \equiv \sum_{\nu'} a\delta_{\nu\nu'}c_{\nu'} \qquad (26.13)$$

where $\delta_{\nu\nu'}$ is a unit matrix in Hilbert space.

With the help of the matrix $\delta_{\nu\nu'}$, the commutation relation for any two operators which yield a number as a result of the commutation, as for example $\hat{p_x}$ and $\hat{x}$, has the form

$$\sum_{\nu'} ((p_x)_{\mu\nu'} (x)_{\nu'\nu} - (x)_{\mu\nu'} (p_x)_{\nu'\nu}) = \frac{h}{i} \delta_{\mu\nu} \qquad (26.14)$$

If the commutation yields a new operator, as in the case of the angular momentum projections, in matrix form it is written as follows:

$$\sum ((M_x)_{\mu\nu'} (M_y)_{\nu'\nu} - (M_y)_{\mu\nu'} (M_x)_{\nu'\nu}) = ih (M_z)_{\mu\nu} \qquad (26.15)$$

The Diagonal Form of a Matrix. Formula (26.4) can be used to represent the transition to the eigenfunction in the variable ν, that is $\psi (\lambda, \nu) \equiv c_\nu$, in the following symmetrical way:

$$\psi (\lambda, \nu) = \int \psi^* (\nu, x) \psi (\lambda, x) \, dx \qquad (26.16)$$

This is the general transformation formula for an independent variable. But it can also be given a different meaning if we put $\nu = \lambda'$. In other words, for $\nu = \lambda'$, Eq. (26.16) undergoes a transition to a set of variables that are themselves eigenvalues of the given operator:

$$\psi (\lambda', \lambda) = \int \psi^* (\lambda', x) \psi (\lambda, x) \, dx \qquad (26.17)$$

But by virtue of the orthogonality of the eigenfunctions of (25.10a), we obtain

$$\psi (\lambda', \lambda) = \delta_{\lambda'\lambda} \qquad (26.18)$$

The eigenfunction of an operator in variables that is a set of its eigenvalues is simply a δ matrix. In these variables the operator itself appears in a very simple form. This can be seen directly from

Eq. (26.6). Substituting the obtained eigenfunction $c_\lambda = \delta_{\lambda\lambda'}$ $(\nu = \lambda)$ into it, we find that a matrix representing an operator in terms of its own variables retains only the terms that occur along the principal diagonal in the array (26.7):

$$(\lambda)_{\lambda\lambda'} = \lambda\delta_{\lambda\lambda'} \qquad (26.19)$$

In other words, it is said that the operator has been reduced to *diagonal form.* For two operators to be reduceable to diagonal form simultaneously, that is, in the same state, they must commutate. In diagonal form two matrices always commutate, since if $\lambda_{\mu\nu} = \lambda\delta_{\mu\nu}$ and $\eta_{\nu\epsilon} = \eta\delta_{\nu\epsilon}$

$$(\lambda\eta)_{\mu\epsilon} = \lambda\eta \sum_\nu \delta_{\mu\nu}\delta_{\nu\epsilon} = \lambda\eta\delta_{\mu\epsilon}$$

is equal to

$$(\eta\lambda)_{\mu\epsilon} = \eta\lambda \sum_\nu \delta_{\nu\epsilon}\delta_{\mu\nu} = \eta\lambda\delta_{\mu\epsilon} = \lambda\eta\delta_{\mu\epsilon}$$

But the commutativity of two matrices cannot depend on the set of variables in which they are expressed. As pointed out, a transition to another set of variables represents a rotation of the coordinate axes in Hilbert space. But if some matrix, in the present case one equal to the commutator of two other matrices, is not equal to zero in one system of axes, no rotation can make it vanish in another system. For two matrices to commutate in one coordinate system they must commutate in all systems. In particular, the respective operators should commutate not only in matrix form but in coordinate representation as well.

Then each of the matrix indices μ, ν, ϵ in the expressions just written corresponds to a certain set of simultaneous eigenvalues of the operators $\hat\lambda$ and $\hat\eta$. This is the difference between these expressions and formula (26.19), which refers only to the eigenvalues of one operator $\hat\lambda$.'

The variables may assume either a discrete or a continuous set of values. For example, rotation angles vary continuously, whereas the eigenvalues of the angular momentum square and its projection vary discretely. It is therefore desirable to give such a meaning to the transformation formulas that would make them equally applicable to a discrete and a continuous set of variables.

Suppose the variable λ varies continuously. It was pointed out in Section 24 that the eigenfunction of an operator in terms of its variable (in this case x) is equal to zero everywhere except at some point $x = x'$ corresponding to the eigenvalue. Hence, if the eigenvalues λ form a continuous set, then

$$\psi(\lambda', \lambda) = \int \psi^*(\lambda', x)\psi(\lambda, x)\,dx = 0 \quad \text{at} \quad \lambda' \neq \lambda \quad (26.20)$$

which agrees with the orthogonality condition (25.6). But if in the case of a discrete set of eigenvalues λ at $\lambda = \lambda'$ the integral (26.20) is equal to unity, for a continuous set the situation changes. To investigate this case, we multiply both sides of (26.20) by some function $c(\lambda)$ and integrate over all allowed values of λ and, furthermore, interchange the integrations over λ and over x on the right to get

$$\int \psi(\lambda',\ \lambda)\, c(\lambda)\, d\lambda = \int \psi^*(\lambda',\ x)\, dx \int c(\lambda)\, \psi(\lambda,\ x)\, d\lambda \quad (26.21)$$

Then the integral over λ on the right, that is $\int c(\lambda)\, \psi(\lambda,\ x)\ d\lambda$, can be treated as a generalization of the series (25.8) for a continuous spectrum of λ:

$$\psi(x) = \int c(\lambda)\, \psi(\lambda,\ x)\, d\lambda \quad (26.22)$$

We now state the basic requirement with respect to functions of a continuous spectrum: the expansion coefficients in the series (25.8) and the integrand $c(\lambda)$ in the expansion (26.22) must be expressed completely analogously, that is (see (25.11a))

$$c(\lambda) = \int \psi^*(\lambda, x)\, \psi(x)\, dx \quad (26.23)$$

But it is apparent from this equation that the expression appearing in the right-hand side of (26.21) is nothing but $c(\lambda')$:

$$\int \psi^*(\lambda,\ x)\, dx \int c(\lambda)\, \psi(\lambda,\ x)\, d\lambda$$
$$= \int \psi^*(\lambda',\ x)\, \psi(x)\, dx = c(\lambda') \quad (26.24)$$

Thus, the eigenfunction $\psi(\lambda',\ \lambda)$ possesses the remarkable property that in integration it simply replaces the argument of the function it multiplies in the integrand:

$$\int \psi(\lambda',\ \lambda)\, c(\lambda)\, d\lambda = c(\lambda') \quad (26.25)$$

If we wish to preserve the analogy with the discrete spectrum of λ, it is convenient to introduce in place of the $\delta_{\lambda\lambda'}$ matrix a similar notation, the δ function. Denoting $\psi(\lambda',\ \lambda) \equiv \delta(\lambda' - \lambda)$, we write

$$\int \delta(\lambda' - \lambda)\, c(\lambda)\, d\lambda = c(\lambda') \quad (26.26)$$

Taking $c(\lambda) = 1$, we find from the preceding equation that

$$\int \delta(\lambda' - \lambda)\, d\lambda = 1 \quad (26.27)$$

We have thus obtained the two basic properties of the δ function introduced by Dirac: it is equal to zero everywhere except the point where its argument is zero, and at that point it becomes infinite in such a way that its integral is equal to unity.

The normalization condition in a continuous spectrum is written with the help of the δ function in the following way, generalizing (26.20):

$$\int \psi^* (\lambda', \, x) \, \psi (\lambda, \, x) \, dx = \delta (\lambda' - \lambda) \qquad (26.28)$$

We thus see that, although the integral of $| \psi (\lambda, \, x) |^2$ becomes infinite, the normalization of functions in a continuous spectrum involves no difficulties. The normalization should be to the δ function, not to unity as in the discrete spectrum.

Differentiating both sides of Eq. (26.26) with respect to λ', we find that

$$\int \left(\frac{d}{d\lambda'} \, \delta (\lambda' - \lambda) \right) c (\lambda) \, d\lambda = \frac{dc (\lambda')}{d\lambda'} \qquad (26.29)$$

If necessary, this formula may be differentiated as many times as there are derivatives of $c (\lambda')$. Thus, the δ function under the integral sign can be differentiated as many times as necessary.

Actually, we have already encountered the δ function in discussing the distribution of a charge (that is, a function of its density) for the case of a point charge. Then, too, the density integral over the volume was equal to a finite quantity, as in formula (26.27).

The Transformation to Momentum Representation. As an example let us examine the form quantum mechanical equations take if the projections rather than the coordinates are taken as the independent variables. Note, firstly, that in coordinate representation, that is, with x, y, z as independent variables, quantum mechanical equations can also be given the general matrix form (26.6). For that the operator $\hat{\lambda} (x)$ should be represented in the form $\hat{\lambda} (x') \, \delta(x - x')$ and integration should be introduced instead of summation. The derivatives involved in the operator $\hat{\lambda} (x)$ are replaced by the corresponding derivatives of the δ function, as in Eq. (26.29): we must go over from $\partial \psi / \partial x$ to $(\partial \psi / \partial x') \, \delta(x' - x)$.

Although it would seem that the coordinate form of writing quantum mechanical equations is simpler than all others, in many cases other representations may offer substantial advantages, which will be made clear from subsequent applications. Werner Heisenberg, who obtained the basic equations of quantum mechanics independently of Schrödinger, from the outset employed explicit matrix representation rather than coordinate representation.

Let us now find the position multiplication operator in momentum representation. For that we interchange λ and ν in (26.16) to get

$$\psi(\nu, \lambda) = \int \psi^*(\lambda, x)\, \psi(\nu, x)\, dx \qquad (26.30)$$

We have obtained the eigenfunction of $\hat{\nu}$ in terms of λ variables. But from a comparison with (26.16) it is apparent that $\psi(\nu, \lambda)$ is the complex conjugate of the function $\psi(\lambda, \nu)$:

$$\psi(\nu, \lambda) = \psi^*(\lambda, \nu) \qquad (26.31)$$

It is apparent from this how the eigenfunction of the position operator is expressed in terms of the momentum variable: if $\psi(p_x, x) = e^{ip_x x/h}$, then $\psi(x, p_x) = e^{-ip_x x/h}$. Since the eigenvalue equation for position is

$$\hat{x}\psi(x, p_x) = x\psi(x, p_x)$$

in the momentum variables we must put[2]

$$\hat{x} = -\frac{h}{i}\frac{\partial}{\partial p_x} \qquad (26.32)$$

In momentum representation the kinetic-energy operator becomes a numerical factor:

$$\hat{T} = \frac{1}{2m}(p_x^2 + p_y^2 + p_z^2) \qquad (26.33)$$

However, the potential energy operator may prove to be more cumbersome than in coordinate representation, that is, it does not have the form of a matrix multiplied by a δ function.

Let us find the potential energy operator for the concrete case of the Coulomb field. For that we must first normalize the eigenfunction of $\psi(p_x, x)$ in a continuous spectrum. Recall the Fourier integral theorem presented in Section 19: if two functions are connected by an integral relationship

$$f(u) = \frac{1}{(2\pi)^{1/2}} \int_{-\infty}^{\infty} e^{iux} g(x)\, dx$$

then the inverse transformation from $f(u)$ to $g(x)$ is written as

$$g(x) = \frac{1}{(2\pi)^{1/2}} \int_{-\infty}^{\infty} e^{-iux} f(u)\, du$$

These equations should be compared with (26.22) and (26.23). In (26.22) we must substitute p_x for λ and the fraction $e^{ip_x x/h}/(2\pi h)^{1/2}$

---

[2] Such a simple transition is possible only in this case.

for $\psi\,(\lambda,\ x)$. It is then apparent that

$$\psi\,(x) = \int\limits_{-\infty}^{\infty} \frac{e^{ip_x x/h}}{(2\pi h)^{1/2}}\, c\,(p_x)\, dp_x, \quad c\,(p_x) = \int\limits_{-\infty}^{\infty} \frac{e^{-ip_x x/h}}{(2\pi h)^{1/2}}\, \psi\,(x)\, dx \tag{26.34}$$

already with the correct normalization, since it was precisely (26.22) and (26.23) that defined the normalization condition.

Thus, to find the matrix element of the Coulomb potential energy we must compute the following integral:

$$\left(\frac{1}{r}\right)_{\mathbf{p}'\mathbf{p}} = \frac{1}{(2\pi h)^3} \int e^{-i\mathbf{p}'\mathbf{r}/h}\, \frac{1}{r}\, e^{i\mathbf{p}\mathbf{r}/h}\, dV \tag{26.35}$$

In it, all components of the momentum vectors $\mathbf{p}'$ and $\mathbf{p}$ appear as each of the indices ν' and ν.

The integral is best found in the following manner. We determine the "potential" φ of a certain "charge" distribution equal to $\rho \equiv$ $\equiv (2\pi h)^{-3}\, e^{i(\mathbf{p}-\mathbf{p}')\mathbf{r}/h}$. In accordance with Section 16, φ satisfies the equation

$$\nabla^2\varphi = -\frac{4\pi}{(2\pi h)^3}\, e^{i(\mathbf{p}-\mathbf{p}')\,\mathbf{r}/h}$$

The potential of a charge element $\rho\, dV$ is $d\varphi = (\rho/r)\, dV$. Then the potential of the whole charge at point $r = 0$ is expressed precisely by formula (26.35). But as can be directly observed, the solution of the potential equation is represented as

$$\varphi = \frac{1}{2\pi^2 h}\, \frac{e^{i\,(\mathbf{p}-\mathbf{p}')\,\mathbf{r}/h}}{(\mathbf{p}-\mathbf{p}')^2} \tag{26.36}$$

which can be verified by substituting (26.36) into the Poisson equation. We obtain expression (26.35) if we substitute $r = 0$. Thus, the matrix element of r^{-1} looks like this:

$$\left(\frac{1}{r}\right)_{\mathbf{p}'\mathbf{p}} = \frac{1}{2\pi^2 h\,(\mathbf{p}-\mathbf{p}')^2} \tag{26.37}$$

and the Schrödinger integral equation for the motion of a particle in a Coulomb field written in terms of momentum variables has the form

$$\frac{\mathbf{p}^2}{2m}\,\psi\,(E,\ \mathbf{p}) - \frac{Ze^2}{2\pi^2 h} \int \frac{\psi\,(E,\ \mathbf{p}')\, d\tau_{\mathbf{p}'}}{(\mathbf{p}-\mathbf{p}')^2} = E\psi\,(E,\ \mathbf{p}) \tag{26.38}$$

In general, if they are expressed in terms of arbitrary variables, all quantum mechanical equations are integral equations. A special feature of the coordinate representation is that it leads to δ' functions for momentum component operators, and the integral equations therefore transform into differential equations. In any case, the

representation of an operator with the help of the δ' function, the derivative of the δ function, cannot be considered diagonal: such a function is other than zero in a domain infinitely close to the zero value of the argument, but never precisely where the argument actually is zero.

The expression (26.32) for the coordinate in terms of the momentum variable is valid, because the momentum and the coordinate are defined in the same interval: they vary continuously from $-\infty$ to ∞. But an angular coordinate—azimuth, for example—varies only between 0 and 2π. This imposes the condition of single-valuedness on the wave function which, as applied to the present case, takes the form

$$\psi(p_\varphi, \varphi) = \psi(p_\varphi, \varphi + 2\pi) \qquad (26.39)$$

Substituting $\psi(p_\varphi, \varphi) = e^{2\pi i p_\varphi \varphi / h}$, we find $p_\varphi = hk$.

We obtained this result before, using the general condition of the single-valuedness of the wave function in its dependence on the spatial coordinates x, y, z, whence it followed that l and k are integers. The example (26.39) was used to show how the restricted nature of the variation interval of the coordinate φ yields a discrete spectrum for the corresponding momentum p_φ. As a consequence, the multiplication operator of angle φ in terms of the variables p_φ does not have a simple form similar to (26.32) (see Exercise 1).

In other words, this can be expressed as follows: the mathematical differential notation of operator p_x (or p_φ) does not effectively define it as long as the boundary conditions imposed on the eigenfunction are not stated. Unlike differential representation, the integral representation of an operator is complete and therefore sometimes preferable.

Unitary Transformations. We shall examine some general properties of transformations from one system of independent variables to another. First, let us find the inverse transformation of (26.16), that is, from the independent variable ν to the variable x:

$$\psi(\lambda, x) = \int \psi^*(x, |\nu) \psi(\lambda, |\nu) \, d\nu \qquad (26.40)$$

The integral form of notation has been used here, but as was shown before, it is in principle no different from notation with the help of a sum for a discrete spectrum. With the help of the relationship (26.31) we eliminate the complex conjugate functions, after which (26.16) and (26.40) take the following form:

$$\psi(\lambda, \nu) = \int \psi(\lambda, x) \psi(x, \nu) \, dx \qquad (26.41a)$$

$$\psi(\lambda, x) = \int \psi(\lambda, \nu) \psi(\nu, x) \, d\nu \qquad (26.41b)$$

The obtained transformations are in form analogous to the matrix multiplication (26.12): it is immaterial whether the indices appear in the subscripts or the arguments of the two variables, or whether we perform a summation or integration with respect to a matrix index.

We can thus say that we have determined the transformation matrix from variables x to variables v:

$$U_{xv} = \psi(x, v) \tag{26.42a}$$

as well as the inverse transformation matrix, which is conventionally written as U_{vx} with the exponent -1:

$$U_{vx}^{-1} = \psi(v, x) \tag{26.42b}$$

But then we see from Eq. (26.31) that between the matrices of the direct and inverse transformations there exists the relationship

$$U_{vx}^{-1} = U_{xv}^* \tag{26.43}$$

The transition to complex conjugate elements with a simultaneous interchange of indices is called *Hermitian conjugation*. A little cross is used instead of an asterisk to distinguish Hermitian conjugates from complex conjugates:

$$U_{vx}^* \equiv U_{xv}^+ \tag{26.44}$$

Applying this notation to the U matrix, we rewrite (26.43) as follows:

$$U_{vx}^{-1} = U_{vx}^+ \tag{26.45}$$

In other words, a matrix that is the inverse of U_{xv} is at the same time *Hermitian conjugate* to it. Such matrices are called *unitary matrices* (we recall that the elements of a Hermitian matrix satisfy the relationship $A_{\alpha\beta}^+ = A_{\alpha\beta}$).

From the definition of an inverse matrix we have

$$\int U^{-1}(v', x) U(x, v)\, dx = \delta_{vv'} \tag{26.46}$$

so that a unitary matrix satisfies the relationship

$$\int U^+(v', x) U(x, v)\, dx = \delta_{vv'} \tag{26.47}$$

which is equivalent to the orthogonality condition of the eigenfunctions of the operator $\hat{v}$ in terms of the variables x.

Transformation of an operator to another representation is also performed with the help of a unitary matrix U. Let us show this on the basis of (26.5). Note, first, that the operator $\hat{\lambda}$ in the integrand is expressed in terms of the variables x, and for that reason alone there is only one integration. In general form operators are expressed

by matrices which, as pointed out, involve δ' functions, making it possible to reduce integration to differentiation only in the x representation. Therefore, the more general notation of (26.5), in which x need not necessarily be interpreted as a coordinate, is

$$\lambda_{v'v} = \int \int \psi^* (v', x') \lambda_{x'x} \psi (v, x) \, dx' \, dx \qquad (26.48)$$

With the help of (26.31), (26.42a), and (26.42b), we give the formula for transforming an operator to new variables of the form

$$\lambda_{v'v} = \int \int U_{v'x'}^{-1} \lambda_{x'x} U_{xv} \, dx \, dx'$$

$$= \int \int U_{v'x'}^{+} \lambda_{x'x} U_{xv} \, dx \, dx' \qquad (26.49)$$

that is, involving the product of three matrices, U^+, λ, and U.

The products of matrices multiplied by vectors or by one another are frequently written without explicitly denoting the indices. Then the transformation formulas from one set of variables to another for wave functions are expressed as

$$\psi' = \psi U \qquad (26.50)$$

and for operators as

$$\lambda' = U^+ \lambda U \qquad (26.51)$$

while the unitary condition is

$$U^+ U = 1$$

where the unity in the right-hand side actually denotes a δ matrix or δ function.

The unitarity of the transformation matrix U is similar to the same condition for the cosines of the rotation angles between old and new Cartesian coordinate axes. If the direct transformations are written as $x_\alpha' = A_{\alpha\beta} x_\beta$, the inverse transformations have the form $x_\beta = A_{\beta\alpha} x_\alpha'$ ($A_{\alpha\beta}$ is the cosine of the angle between the new axis with the label α and the old axis with the label β). But inverse transformations can be expressed with the help of the inverse matrix $A_{\alpha\beta}^{-1}$, so that $A_{\alpha\beta}^{-1} = A_{\beta\alpha}$. In Hilbert space, where vectors are complex, complex conjugation is performed in addition to interchanging the indices.

Determination of the eigenvalues of an operator represents in effect the reduction of a second-order "surface" in Hilbert space to the principal axes. Two surfaces cannot be reduced to the principal axes simultaneously because the respective principal axes of these surfaces are differently oriented. Such is the geometric interpretation

of the provision that two operators do not have eigenvalues in the same state, that is, the corresponding quantities cannot exist simultaneously.

EXERCISES

1. Find the matrix elements of the multiplication operator of the rotation angle φ in terms of the variables $p_\varphi = hk$.

Answer. From Eq. (26.5) we obtain

$$\varphi_{k'k} = \frac{i}{k' - k} \quad \text{for} \quad k' \neq k$$

$$= \pi \quad \text{for} \quad k' = k$$

2. Show that a unitary operator U can be represented in the form

$$U = e^{i\Phi}$$

where Φ is a certain Hermitian operator.

Solution. If

$$U = e^{i\Phi}$$

then

$$U^{-1} = e^{-i\Phi}$$

Let us now show that $U^{-1} = U^+$. For this, note, firstly, that a transformation to a complex conjugate operator in any case means a reversal of the sign of Φ in the exponent. Further, we expand the exponent in a series:

$$U^{-1} = 1 - i\Phi + \frac{i^2}{2}\Phi^2 - \frac{i3}{6}\Phi^3 + \cdots$$

We now take one of the terms of the series, the third, for example, and write it with the help of matrix indices, agreeing that we make use of the summation condition and know the Hermiticity of Φ^+ ($\Phi^+_{\alpha\beta} = \Phi_{\alpha\beta}$):

$$\Phi^3_{\alpha\beta} = \Phi_{\alpha\mu}\Phi_{\mu\nu}\Phi_{\nu\beta} = \Phi^+_{\alpha\mu}\Phi^+_{\mu\nu}\Phi^+_{\nu\beta} = (\Phi^+)^3_{\alpha\beta}$$

Collecting the expansion terms in exponential form, we obtain

$$U^{-1} = U^+$$

27

OPERATORS IN MATRIX REPRESENTATION

The Time Dependence of Matrix Elements. The results obtained in the preceding sections make it possible to find the solutions of certain important quantum mechanical problems. Let us first show that it is possible to write equations of motion for operators similar to the equations of motion of Newtonian mechanics.

We write the matrix element of a certain operator in terms of energy variables, that is

$$\lambda_{E'E} = \int \psi^* (E', x) \hat{\lambda} \psi (E, x) \, dx \qquad (27.1)$$

The eigenfunctions of the Hamiltonian in the integrand, together with the time dependent factor, have the form (23.20):

$$\psi(E, x) = e^{-iEt/\hbar} \psi_0 (E, x) \qquad (27.2)$$

where the zero subscript of the wave function denotes that the time factor has been separated. Then the matrix element (27.1) takes the form

$$\lambda_{E'E} = e^{-i(E-E')t/\hbar} \int \psi_0^* (E', x) \hat{\lambda} \psi_0 (E, x) \, dx \qquad (27.3)$$

A matrix element without the time factor is said to be written in the *Schrödinger representation*, and with the factor, in *Heisenberg representation*.

We now differentiate the matrix element with respect to time, assuming for generality that $\hat{\lambda}$ may also involve an explicit time dependence:

$$\frac{d\lambda_{E'E}}{dt} = e^{-i(E-E')t/\hbar}$$

$$\times \int \psi_0^* (E', x) \left(\frac{\partial \hat{\lambda}}{\partial t} - \frac{i}{\hbar} (E - E') \hat{\lambda} \right) \psi_0 (E, x) \, dx \qquad (27.4)$$

Returning to the matrix element (27.1), we rewrite Eq. (27.4) as follows:

$$\frac{d\lambda_{E'E}}{dt} = \frac{\partial \lambda_{E'E}}{\partial t} + \frac{i}{\hbar} (E'\lambda_{E'E} - \lambda_{E'E}E) \qquad (27.5a)$$

We know from the preceding section (see Eq. (26.19)) that an operator written in terms of its own variables has diagonal form. Hence

$$\mathcal{H}_{E'E} = E'\delta_{E'E'}, \qquad \mathcal{H}_{E'E} = \delta_{E'E}E \qquad (27.6)$$

But then (27.5*a*) can be written using the commutator for the Hamiltonian $\hat{\mathscr{H}}$ and the operator $\hat{\lambda}$:

$$\frac{d\lambda_{E'E}}{dt} = \frac{\partial \lambda_{E'E}}{\partial t} + \frac{i}{h}(E'\delta_{E'E''}\lambda_{E''E} - \lambda_{E'E''}\delta_{E''E}E)$$

$$= \frac{\partial \lambda_{E'E}}{\partial t} + \frac{i}{h}((\mathscr{H}\lambda)_{E'E} - (\lambda\mathscr{H})_{E'E}) \qquad (27.7)$$

A matrix equation holds in any system of variables, notably in coordinate form:

$$\frac{d\hat{\lambda}}{dt} = \frac{\partial \hat{\lambda}}{\partial t} + \frac{i}{h}(\hat{\mathscr{H}}\hat{\lambda} - \hat{\lambda}\hat{\mathscr{H}}) \qquad (27.8)$$

Note that Eq. (27.8) has a classical analogue. If there is a certain dynamical quantity λ (p, q; t), its total time derivative is

$$\frac{d\lambda}{dt} = \frac{\partial \lambda}{\partial t} + \frac{\partial \lambda}{\partial p}\frac{dp}{dt} + \frac{\partial \lambda}{\partial q}\frac{dq}{dt} \qquad (27.9)$$

Substituting $\dot{p}$ and $\dot{q}$ from Hamilton's equations (10.6*a*) and (10.6*b*), we obtain

$$\frac{d\lambda}{dt} = \frac{\partial \lambda}{\partial t} + \frac{\partial \lambda}{\partial q}\frac{\partial \mathscr{H}}{\partial p} - \frac{\partial \lambda}{\partial p}\frac{\partial \mathscr{H}}{\partial q} \qquad (27.10)$$

It will be shown in Section 31 that the expression (27.10) develops from (27.8) in the limiting transition to classical mechanics. The expression $\frac{\partial \lambda}{\partial q}\frac{\partial \mathscr{H}}{\partial p} - \frac{\partial \lambda}{\partial p}\frac{\partial \mathscr{H}}{\partial q}$ is called the *Poisson bracket*.

If the quantity λ does not explicitly depend on time and its Poisson bracket is not zero, then $\dot{\lambda} = 0$, that is λ is an integral of the motion. Similarly, if the operator $\hat{\lambda}$ does not involve t explicitly and is commutative with the Hamiltonian, then it is called the *quantum integral of the motion*. The corresponding quantity λ exists in the given state simultaneously with the energy of the system.

For example, the Hamiltonian of a particle in a central Coulomb field has the form (see (24.32))

$$\hat{\mathscr{H}} = -\frac{h^2}{2m}\frac{1}{r^2}\frac{\partial}{\partial r}r^2\frac{\partial}{\partial r} + \frac{\hat{M}^2}{2mr^2} - \frac{Ze^2}{r} \qquad (27.11)$$

The operator of the angular momentum square involves only differentiation with respect to the angles and therefore commutes with the Hamiltonian. Hence, it is conserved together with the energy. The same holds for the operator of the angular momentum projection $\hat{M}_z = (h/i)(\partial/\partial\varphi)$. Thus, within the framework permitted by quantum mechanics, the same quantities are conserved together with the energy as in classical mechanics, where all three components of angular momentum were conserved. Only in quantum mechanics,

instead of the three components, we have the square of the angular momentum and its projection. Conservation of the three angular momentum projections corresponds to a path of motion lying in a plane perpendicular to the angular momentum, but in quantum mechanics there are no paths.

Let us now find the quantum analogue of the equations of motion, that is, calculate the total time derivatives of the coordinate and momentum. We have

$$\frac{d\hat{\mathbf{r}}}{dt} = \frac{i}{h}(\hat{\mathscr{H}}\hat{\mathbf{r}} - \hat{\mathbf{r}}\hat{\mathscr{H}}) = \frac{i}{h}\left(\frac{\hat{p}^2}{2m}\hat{\mathbf{r}} - \hat{\mathbf{r}}\frac{\hat{p}^2}{2m}\right) \qquad (27.12)$$

because the potential energy operator depends only upon the coordinate and commutes with $\hat{\mathbf{r}}$. To find the obtained commutation relation, let us study it for one component of $\hat{\mathbf{r}}$, for example, x. For this component

$$\hat{p}^2\hat{x} - \hat{x}\hat{p}^2 = \hat{p}_x^2\hat{x} - \hat{x}\hat{p}_x^2$$

since the squares of the other two momentum components commute with x. Further, we add to and subtract from the commutator the same term $\hat{p}_x\hat{x}\hat{p}_x$. Collecting like terms, we obtain

$$\hat{p}_x(\hat{p}_x\hat{x} - \hat{x}\hat{p}_x) + (\hat{p}_x\hat{x} - \hat{x}\hat{p}_x)\,\hat{p}_x = 2\frac{h}{i}\,\hat{p}_x$$

Commutation relations for the other components of the radius vector are found similarly. Substituting them into (27.12), we finally get

$$\hat{\dot{x}} = \frac{\hat{p}_x}{m}, \qquad \hat{\dot{\mathbf{r}}} = \frac{\hat{\mathbf{p}}}{m} \qquad (27.13)$$

Hence, the velocity and momentum operators are connected by the same relationship as classical quantities.

Let us determine the derivative of the momentum operator:

$$\hat{\dot{p}}_x = -\frac{\partial \hat{U}}{\partial x}, \qquad \hat{\dot{\mathbf{p}}} = -\frac{\partial \hat{U}}{\partial \mathbf{r}} \qquad (27.14)$$

where the result of Exercise 2, Section 24, has been used.

It is natural to call the operator in the right-hand side of Eqs. (27.14) the "force" operator. Thus, we have obtained for operators the same equations of motion, (27.13) and (27.14), as for classical quantities. This assertion is known as the *correspondence principle*.

Forming mean values of Eqs. (27.13) and (27.14) according to the rule (25.19), we find that quantum mechanical means satisfy the classical equations of motion.

The Linear Harmonic Oscillator. Let us apply Eqs. (27.13) and (27.14) to the problem on determining the energy eigenvalues of a linear harmonic oscillator. We know from Eq. (7.31) that the energy of a separate linear harmonic oscillator of unit mass is

$$E = \frac{\dot{Q}^2}{2} + \frac{\omega^2 Q^2}{2}$$

Suppose the mass of the oscillator is m. Then its x coordinate can be measured in conventional length units, and its momentum p in g-cm-s^{-1} units.

Expressing the velocity in terms of the momentum according to the formula $\dot{x} = p/m$, we represent the Hamiltonian of the oscillator in the form

$$\mathscr{H} = \frac{p^2}{2m} + \frac{m\omega^2 x^2}{2} \tag{27.15}$$

In order to determine the energy eigenvalues, (27.15) must be treated as an operator equation:

$$\hat{\mathscr{H}} = \frac{\hat{p}^2}{2m} + \frac{m\omega^2 \hat{x}^2}{2} \tag{27.16}$$

In this section, we shall solve the problem with energy as the independent variable. At first glance this may appear to be much more difficult than the coordinate representation, where we at least know the form of the operators: $p = (h/i)\,(\partial/\partial x)$. Actually though, in the case of an oscillator the matrix form possesses a number of advantages, since it enables an algebraic solution. We shall solve the oscillator problem in coordinate representation in the next section; this will require an analysis of the differential equation.

We average equation (27.16) over an arbitrary state of the oscillator:

$$\langle E \rangle = \frac{\langle p^2 \rangle}{2m} + \frac{m\omega^2 \langle x^2 \rangle}{2} \tag{27.17}$$

It is apparent that the mean values of the positive quantities $\langle p^2 \rangle$ and $\langle x^2 \rangle$ are at least not negative, so that the mean value of the energy is not negative either. If the averaging is performed over the eigenstate of the Hamiltonian, then the mean value equals the eigenvalue, therefore the energy eigenvalues are not negative.

We now write the equations of motion, (27.13) and (27.14), for the oscillator, that is, putting the potential energy $U = m\omega^2 x^2/2$. This yields

$$\dot{\hat{x}} = \hat{p}/m \tag{27.18}$$

$$\dot{\hat{p}} = -m\omega^2 \hat{x} \tag{27.19}$$

We take the matrix elements, in energy representation, of both sides of these equations, that is, we replace each operator $\hat{\lambda}$ appearing in the equation with its matrix element $\lambda_{E'E}$. Since none of the operators in our problem explicitly depend upon time, we obtain from (27.5)

$$\dot{\lambda}_{E'E} = \frac{i}{h}(E' - E)\,\lambda_{E'E} \tag{27.5b}$$

Applying this formula to (27.18) and (27.19), we find

$$\frac{i}{h}(E' - E)\,x_{E'E} = \frac{1}{m}\,p_{E'E} \tag{27.20}$$

$$\frac{i}{h}(E' - E)\,p_{E'E} = -m\omega^2 x_{E'E} \tag{27.21}$$

We solve the second equation for $p_{E'E}$ and substitute it into the first to get

$$((E' - E)^2 - h^2\omega^2)\,x_{E'E} = 0 \tag{27.22}$$

Thus only such a matrix element $x_{E'E}$ is other than zero for which $E' - E = \pm h\omega$. We see from this that the differences between neighbouring energy eigenvalues can be equal only to $\pm h\omega$ and no other value. On the other hand, we have established that the energy eigenvalues are not negative, so that by subtracting the quantity $h\omega$ from some energy eigenvalue a sufficient number of times we inevitably arrive at some least energy eigenvalue E_0. Therefore, in future we shall write the energy eigenvalue in the form

$$E_n = h\omega n + E_0 \tag{27.23}$$

Accordingly, instead of $x_{E'E}$, we denote the matrix elements $x_{n'n}$, where n' and n are positive integers. In this notation, in place of Eq. (27.22) we write a similar equation in which the labels are not the energy values themselves but their numbers:

$$((n' - n)^2 - 1)\,x_{n'n} = 0 \tag{27.24}$$

First of all, it is apparent that when the numbers are equal ($n' = n$), $x_{n'n} = 0$; hence the matrix $x_{n'n}$ has no diagonal elements. We put $n' = n \pm 1$. Then the first term on the left in Eq. (27.24) vanishes, whence it follows that $x_{n\pm1\,n} \neq 0$. At all other n' this factor is not zero, so that all the other matrix elements are, like the diagonal elements, zero; thus, only $x_{n\pm1\,n} \neq 0$.

Let us now determine the nonzero matrix elements $x_{n\pm1\,n}$. For that we proceed from the commutation relation (24.17), which should also be rewritten in energy representation. In the right-hand side of Eq. (24.17) we have a number, so that only diagonal matrix elements result from the right-hand, and correspondingly left-hand,

sides. In order to determine the matrix elements of the left-hand side we must express the matrix elements of momentum in terms of the coordinate matrix elements. This is done with the help of the equation of motion (27.20), which shows that nonzero elements in $p_{n'n}$ have the same labels as in matrix $x_{n'n}$, that is $p_{n\pm1\ n}$. For these elements we obtain

$$p_{n+1\ n} = im\omega x_{n+1\ n}, \qquad p_{n-1\ n} = -im\omega x_{n-1\ n} \qquad (27.25)$$

We find the commutation relations for the left-hand side according to the general rules of matrix multiplication:

$$p_{n\ n+1}x_{n+1\ n} + p_{n\ n-1}x_{n-1\ n} - x_{n\ n+1}p_{n+1\ n} - x_{n\ n-1}p_{n-1\ n}$$

$$= -2m\omega i\,|\,x_{n\ n+1}\,|^2 + 2m\omega i\,|\,x_{n\ n-1}\,|^2 = \frac{h}{i} \qquad (27.26a)$$

But since $x_{n\ n-1}$ is Hermitian, $x_{n\ n-1} = x^*_{n-1\ n}$, and therefore $|\,x_{n\ n-1}\,|^2 = |\,x_{n-1\ n}\,|^2$. Equation (27.26a) can also be rewritten in the form

$$|\,x_{n\ n+1}\,|^2 - |\,x_{n-1\ n}\,|^2 = \frac{h}{2m\omega} \qquad (27.26b)$$

We see that each label in the first term on the left is greater by unity than the corresponding label in the second term. Furthermore, there are no terms with negative labels, since according to (27.23) n begins at zero. Therefore

$$|\,x_{01}\,|^2 = \frac{h}{2m\omega}, \qquad |\,x_{12}\,|^2 = 2\,\frac{h}{2m\omega}, \ \ldots$$

$$|\,x_{n\ n+1}\,|^2 = (n+1)\,\frac{h}{2m\omega} \qquad (27.27)$$

Since $|\,x_{n\ n+1}\,|^2 = |\,x_{n+1\ n}\,|^2$, we have determined the squares of the moduli of all nonzero matrix elements of the coordinate.

We assume that the matrix elements themselves are real numbers and the value of their phase is not involved in any physical equation. Then, from (27.25), the matrix elements of the momentum are purely imaginary numbers. It follows then from the Hermiticity of the momentum operator that $p_{n'n} = p^*_{nn'} = -p_{nn'}$. Making use of (27.25) and (27.27), we write both matrices, $x_{n'n}$ and $p_{n'n}$, explicitly

$$x_{n'n} = \left(\frac{h}{2m\omega}\right)^{1/2} \begin{pmatrix} 0 & \sqrt{1} & 0 & 0 & \ldots \\ \sqrt{1} & 0 & \sqrt{2} & 0 & \ldots \\ 0 & \sqrt{2} & 0 & \sqrt{3} & \ldots \\ 0 & 0 & \sqrt{3} & 0 & \ldots \\ & & \ldots & & \end{pmatrix}$$

$$p_{n'n} = i\left(\frac{mh\omega}{2}\right)^{1/2}\begin{pmatrix} 0 & -\sqrt{1} & 0 & 0 & \ldots \\ \sqrt{1} & 0 & -\sqrt{2} & 0 & \ldots \\ 0 & \sqrt{2} & 0 & -\sqrt{3} & \ldots \\ 0 & 0 & \sqrt{3} & 0 & \ldots \\ \ldots & \ldots & \ldots & \ldots & \ldots \end{pmatrix} \quad (27.28a)$$

It remains for us to compute the lowest energy E_0 (the energy of the ground state). It can be represented as the diagonal matrix element of $\hat{\mathscr{H}}$ with labels 0, 0. Expressing the energy matrix in terms of the squares of the coordinate and momentum matrices, we obtain

$$E_0 = \mathscr{H}_{00} = \frac{p_{01}p_{10}}{2m} + \frac{m\omega^2 x_{01}x_{10}}{2} = \frac{h\omega}{4} + \frac{h\omega}{4} = \frac{h\omega}{2} \quad (27.28b)$$

where we made use of (27.28a). The meaning of *zero energy* will be explained in the following section.

The Density Matrix. In discussing the Stern-Gerlach experiment in Section 24, we pointed out that a primary beam of particles arbitrarily oriented in a magnetic field splits into $2l + 1$ beams, whereas a beam with a definite value of the angular momentum projection splits only in a field that is not parallel to the z axis, that is, not parallel to the field which caused the primary splitting. In the latter case it is said that a system is in a *pure quantum state*[3] characterized by a definite wave function. (An experiment performed on a system in pure state yields the same result when repeated.)

How is one to characterize the state of the initial beam? We know the probability w_k of the occurrence of every value of k: it is equal to $(2l + 1)^{-1}$. In a pure state we must know not only the probabilities w_k but also their amplitudes c_k, in terms of which w_k is expressed as $|c_k|^2$. Hence, in a pure state a system may be characterized by a wave function

$$\psi(x) = \sum_k c_k \psi(k, x)$$

But for this the system must be closed, that is, it should not interact with anything. Only in that case does it have a definite Hamiltonian and its wave function $\psi(x)$ satisfies the exact wave equation.

In the Stern-Gerlach experiment, the particles (usually atoms of a metal) are produced by evaporation of a substance in a special oven. In such conditions an atom cannot be treated as a closed system, since it interacts with its surroundings. This interaction

---

[3] In our case a separate particle serves as the system.

is not strong enough to affect the angular momentum square of each atom, so that the value $M^2 = h^2 l \, (l + 1)$ is the same for all the atoms (see Sec. 24); however, it creates conditions in which all values of k for the ejected atoms are equiprobable. Such a state, which has been subjected to an external ¦action, is known as a *mixture*, as distinct from the pure state.

We shall not consider mixtures specifically in connection with the Stern-Gerlach experiment, but in the most general case. Let the probability of the occurrence of the nth state of a system as a result of some external action be w_n. To describe such a system it is convenient to introduce a special matrix

$$\rho \, (x', \, x) = \sum_n w_n \psi^* \, (n, \, x') \, \psi \, (n, \, x) \tag{27.29}$$

called the *density matrix* of the system.

Let us show how the mean values are calculated for a given mixture with the help of the density matrix. The general definition of the mean of a certain quantity, $\langle \lambda \rangle$, is, as usual,

$$\langle \lambda \rangle = \sum_n w_n \, \langle \lambda \rangle_n \tag{27.30}$$

where $\langle \lambda \rangle_n$ is the mean of λ over the nth state (cf. (25.18)). From (25.19) the mean value of λ over the nth pure state is

$$\langle \lambda \rangle_n = \int \psi^* \, (n, \, x) \, \hat{\lambda} \psi \, (n, \, x) \, dx \tag{27.31}$$

We substitute this into (27.30) and interchange the order of summation and integration. For this it is first convenient to represent $\hat{\lambda} \, (x)$ in matrix form:

$$\hat{\lambda} \, (x) = \delta \, (x - x') \, \hat{\lambda} \, (x') \equiv \lambda \, (x', \, x) \tag{27.32}$$

as was done in the preceding section. We recall that if $\hat{\lambda} \, (x')$ involves a differentiation with respect to x', then (27.32) is not written in diagonal form. Now, substituting (27.32) into the expression for the mean, we reduce it to the form

$$\langle \lambda \rangle = \int dx \int dx' \sum_n w_n \psi^* \, (n, \, x') \, \lambda \, (x, \, x') \, \psi \, (n, \, x)$$

$$= \int dx \int dx' \lambda \, (x, \, x') \, \rho \, (x', \, x) \tag{27.33}$$

in accordance with (27.29).

This was the form in which we obtained the diagonal element of the product matrix $\lambda \rho$, that is $\int dx' \, \lambda \, (x, \, x') \, \rho \, (x', \, x)$, integrated over x. The sum of the diagonal elements of a matrix, $A_{\nu\nu}$, or the

integral over the diagonal elements $A(x,\ x)$, if the index varies continuously, is called the *trace* (Tr) of the matrix:

$$\operatorname{Tr} A \equiv \sum_{\nu} A_{\nu\nu} \qquad (27.34)$$

$$\operatorname{Tr} A \equiv \int dx\ A(x,\ x) \qquad (27.35)$$

Expression (27.33) is then the trace of the product of the density matrix and the matrix of the averaged quantity:

$$\langle \lambda \rangle = \operatorname{Tr}(\lambda \rho) \qquad (27.36)$$

The significance of Eq. (27.36) is that it does not depend on the adopted coordinate representation. The value of the trace of a matrix is invariant under a unitary transformation, that is, with respect to a transformation to another representation. Let us show this with the help of the equations of the preceding section. We write Eq. (26.48) for an arbitrary operator $\hat{A}$:

$$A_{\nu'\nu} = \int\int \psi^*(\nu',\ x')\, A_{x'x} \psi(\nu,\ x)\, dx'\, dx \qquad (27.37)$$

and find the trace of both sides, interchanging on the right the integrations over $x',\ x$ and the summation over ν:

$$\operatorname{Tr} A = \sum_{\nu} A_{\nu\nu} = \int\int dx'\, dx A_{x'x} \sum_{\nu} \psi^*(\nu,\ x')\, \psi(\nu,\ x) \qquad (27.38)$$

We now take advantage of the fact that $\psi^*(\nu,\ x') = \psi(x',\ \nu)$ and $\psi(\nu,\ x) = \psi^*(x,\ \nu)$. Then the inner sum reduces to the form

$$\sum_{\nu} \psi^*(x,\ \nu)\, \psi(x',\ \nu) \qquad (27.39)$$

But by virtue of the general condition of the orthogonality of wave functions this sum is

$$\sum_{\nu} \psi^*(x,\ \nu)\, \psi(x',\ \nu) = \delta(x - x') \qquad (27.40)$$

Here Eq. (26.28) was used, with the substitution of x for λ and ν for x. If the spectrum of ν is discrete, the integration is replaced by a summation, but if only the spectrum of λ (that is x) is continuous, the right-hand side of the equation is expressed in terms of a δ function. With the help of (27.40) we obtain

$$\operatorname{Tr} A = \sum_{\nu} A_{\nu\nu} = \int dx \int dx'\, A_{xx'} \delta(x - x') = \int dx\, A_{xx} \qquad (27.41)$$

where the fundamental property of the δ function was used.

Thus, the expression of a mean quantity in terms of the trace of the density matrix is valid in any representation, as it should be for any physical quantity.

Incidentally, we note an important property of the trace of a product matrix:

$$\sum_{v} (AB)_{vv} = \sum_{v\mu} A_{v\mu} B_{\mu v} = \sum_{\mu v} A_{\mu v} B_{v\mu}$$
$$= \sum_{\mu v} B_{v\mu} A_{\mu v} = \sum_{v} (BA)_{vv} \qquad (27.42a)$$

(We have, as many times before, interchanged the indices with respect to which the summation was performed.) In abbreviated form we write

$$\text{Tr } (AB) = \text{Tr } (BA) \qquad (27.42b)$$

The invariance of the trace in Hilbert space has an analogue in Euclidean space: the sum of the diagonal components of a tensor of rank 2, $A_{\alpha\alpha}$, is a scalar, and it does not change in a rotation of the coordinate axes.

With the help of the fundamental equation (27.36) it is simple to find the trace of the density matrix itself. For that it is sufficient to substitute $\lambda = 1$ into the formula. We obtain

$$\text{Tr } \rho = 1 \qquad (27.43)$$

We now find the trace of the square of the density matrix, that is, the product of a multiplication of ρ by itself. According to the rule of matrix multiplication,

$$(\rho \times \rho)_{xx'} = \int dx'' \rho (x, x'') \rho (x'', x') \qquad (27.44)$$

We substitute the expressions for the matrices $\rho(x', x'')$ and $\rho(x'', x)$ and interchange the summation and integration order. Making use of the orthogonality of the functions $\psi (n, x'')$ and $\psi^* (n', x'')$, we obtain

$$\int dx'' \rho (x, x'') \rho (x'', x')$$
$$= \sum_{n, n'} w_n w_{n'} \psi^* (n, x) \psi (n', x') \int \psi^* (n', x'') \psi (n, x'') dx''$$
$$= \sum_{n} w_n^2 \psi^* (n', x) \psi (n, x)$$

To find the trace of this expression we must put $x' = x$ and integrate over x. Thanks to normalization, every integral of $| \psi (n, x)|^2$

is equal to unity, so that

$$\text{Tr}\,(\rho \times \rho) = \sum_n w_n^2 \qquad (27.45)$$

But the trace of the density matrix is, as we already know from (27.43),

$$\text{Tr}\,\rho = \sum_n w_n \int \psi^*\,(n,\,x)\,\psi\,(n,\,x)\,dx = \sum_n w_n = 1 \qquad (27.46)$$

If the probability of some event does not become a certainty, then $w_n < 1$. Hence, the square of the probability is smaller than the probability itself: $w_n^2 < w_n$. Comparing (27.45) with (27.46), we see that

$$\text{Tr}\,(\rho \times \rho) \leqslant \text{Tr}\,\rho \qquad (27.47a)$$

The inequality may become an equality only in the case of a pure state, when one of the probabilities, for example w_n, is equal to unity, and all $w_{n'\neq n} = 0$. Then

$$\text{Tr}\,(\rho \times \rho) = \text{Tr}\,\rho = 1 \qquad (27.47b)$$

The degree to which $\text{Tr}\,(\rho \times \rho)$ approximates to $\text{Tr}\,\rho$ can be treated as a measure of the purity of the state.

Thus, the density matrix provides the fullest possible description of a mixture.

Let us now find the *equation of motion of the density matrix*, which replaces the Schrödinger equation (23.11), $-(h/i)\,(\partial\psi/\partial t) = \hat{\mathcal{H}}\psi$. We take the derivative of the density matrix with respect to time:

$$\frac{h}{i}\,\frac{\partial\rho\,(x',\,x)}{\partial t}$$

$$= \sum w_n \left(\frac{h}{i}\,\frac{\partial\psi^*\,(n,\,x')}{\partial t}\,\psi\,(n,\,x) + \psi^*\,(n,\,x')\,\frac{h}{i}\,\frac{\partial\psi\,(n,\,x)}{\partial t} \right)$$

But

$$\frac{h}{i}\,\frac{\partial\psi^*}{\partial t} = \hat{\mathcal{H}}^*\psi^*, \qquad \frac{h}{i}\,\frac{\partial\psi}{\partial t} = -\,\hat{\mathcal{H}}\psi$$

whence we obtain

$$\frac{\partial\rho\,(x',\,x)}{\partial t} = \frac{i}{h}\,(\hat{\mathcal{H}}^*\,(x')\,\rho\,(x',\,x) - \hat{\mathcal{H}}\,(x)\,\rho\,(x',\,x)) \qquad (27.48)$$

Since this equation is in operator form, it is valid in any representation.

28

SOME PROBLEMS
IN COORDINATE REPRESENTATION

In this section we shall obtain solutions of the wave equation for certain cases, which are in part illustrative and in part of an auxiliary nature. Nevertheless, they help to reveal many important regularities.

Problems involving boundary conditions imposed on the wave function in its dependence on the coordinates would be difficult to solve in any but the coordinate representation. The greater part of this section is devoted to such problems. In addition, a solution is presented of the linear harmonic oscillator problem with the help of the Schrödinger equation. Irrespective of the earlier obtained result in the energy representation, the solution of this problem in the coordinate representation is of major interest as an illustration of the computational methods of quantum mechanics.

A Particle in a One-Dimensional, Infinitely Deep Potential Well.
Suppose a particle is constrained to move in one dimension remaining in an interval of length a, so that $0 \leqslant x \leqslant a$. We can imagine
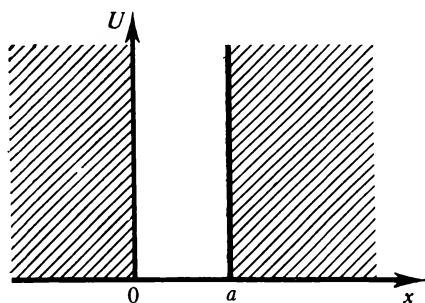
Figure 30

that at points $x = 0$ and $x = a$ there are absolutely impenetrable walls which reflect the particle. A limitation of this type is represented with the aid of the potential energy curve shown in Figure 30: $U = \infty$ at $x < 0$ and at $x > a$. We put $U = 0$ at $0 \leqslant x \leqslant a$; this is the potential energy gauge. To leave the region $0 \leqslant x \leqslant a$, a particle would have to perform an infinitely large quantity of work. Thus, the probability for the particle to be at $x = 0$ or $x = a$

is equal to zero. With the aid of (23.1), we obtain

$$\psi(0) = \psi(a) = 0 \tag{28.1}$$

These boundary conditions may also be justified by means of a limiting transition from a well of finite depth to a well of infinite depth. This will be done later.

Insofar as the potential energy is time independent, we can write the equation for the energy eigenvalues directly; this is Eq. (23.21). The motion is one-dimensional and, therefore, we must take the total derivative d^2/dx^2 in place of ∇^2. From this we have

$$-\frac{h^2}{2m}\frac{d^2\psi}{dx^2} = E\psi \tag{28.2}$$

We introduce the shortened notation

$$\frac{2mE}{h^2} = \varkappa^2 \tag{28.3}$$

so that the wave equation will be of the form

$$\frac{d^2\psi}{dx^2} = -\varkappa^2\psi \tag{28.4}$$

The solution to (28.4) is well known:

$$\psi = C_1 \sin \varkappa x + C_2 \cos \varkappa x \tag{28.5}$$

But from (28.1) $\psi(0) = 0$, so that the cosine term must be omitted by putting $C_2 = 0$. There remains

$$\psi = C_1 \sin \varkappa x \tag{28.6}$$

We now substitute the second boundary condition

$$\psi(a) = C_1 \sin \varkappa a = 0 \tag{28.7}$$

This is an equation in $\varkappa$. It has an infinite number of solutions:

$$\varkappa a = n\pi \tag{28.8}$$

where n is any nonzero integer:

$$1 \leqslant n \leqslant \infty \tag{28.9}$$

The value $n = 0$ is precluded, because at $n = 0$ the wave function vanishes everywhere ($\psi = \sin 0 = 0$); hence $|\psi|^2 = 0$, and the particle simply does not exist anywhere (a trivial solution).

Now substituting $\varkappa$ from the definition (28.3) and solving (28.8) with respect to E, we find the expression for energy, that is, the

energy spectrum for the problem being examined:

$$E_n = \frac{\pi^2 h^2}{2ma^2} n^2 \tag{28.10}$$

The boundary condition imposed on a wave function is just as essential for finding the energy spectrum as the wave equation itself. As is apparent from (28.10), it is valid not for all energy values but only for those that belong to a definite set of numbers characteristic of the given problem. Depending on the conditions, these numbers may form either a discrete series, as in the present problem, or a continuous set, as in the problem on the free motion of a particle.

Indeed, in the free motion of a particle its wave function must remain finite everywhere. This condition is satisfied by the function (23.3) at all real and positive energy values. In this case imaginary momentum would correspond to negative energies, and the coordinate dependence of the wave function would have the form of an exponential with a real exponent. But such a function becomes infinite at $x = +\infty$ or $x = -\infty$.

The solution of Eq. (23.21) for stationary states is always associated with finding the energy spectrum. In contrast with Bohr's theory, where the discreteness of states was a necessary, but alien, appendage to classical motion, in quantum mechanics the very character of the motion determines the energy spectrum. This will be made especially apparent in the examples that follow.

Let us now return to the wave function (28.6). It vanishes within the interval $(0, a)$ (that is, except at its ends) n times. The number of zeros (nodes) of the wave function equals the number of the energy eigenvalue.

This result is easily understood from the following reasoning. At $n = 1$, there is one sinusoidal half-wave in the interval $(0, a)$; at $n = 2$, there is one wave; at $n = 3$ there are three half-waves, etc. Hence, the greater the value of n the smaller the de Broglie wavelength λ. But energy is proportional to the square of momentum, that is, inversely proportional to the square of λ, according to (22.2a). Hence, the smaller the λ the greater the energy. This conclusion holds, of course, not only for wave functions of a purely sinusoidal shape, though as a qualitative rather than an exact quantitative relationship: the more zeros, or nodes, the wave function has, the greater the energy.

The least-energy state corresponds to a wave function which has no nodes anywhere within the interval. It is called the *ground state*, all the other states being termed *excited*.

It remains to determine the coefficient C_1 in order to define the wave function completely. We shall find it from the normalization

condition (23.17):

$$\int_0^a |\psi|^2\, dx = C_1^2 \int_0^a \sin^2 \varkappa x\, dx = C_1^2 \int_0^a \frac{1 - \cos 2\varkappa x}{2}\, dx$$

$$= C_1^2 \left(\frac{x}{2} - \frac{\sin 2\varkappa x}{4\varkappa} \right) \Big|_0^a = \frac{C_1^2 a}{2} = 1$$

The second term of the integrated expression becomes zero at both limits in accordance with (28.8). Thus

$$C_1 = (2/a)^{1/2} \tag{28.11}$$

$$\psi_n = \left(\frac{2}{a} \right)^{1/2} \sin \frac{\pi n x}{a} \tag{28.12}$$

The wave function (28.12) is real. Therefore, from (23.19) the current (particle flux) in this state is zero. This can also be explained as follows. The wave function (28.12) separates into a sum of two exponentials. Together with the time factor $e^{-iEt/h}$, each such exponential represents the wave function of a free particle, (23.3), one of them corresponding to momentum $p = h\varkappa$, and the other to the same momentum but with opposite sign. Thus, state (28.12) represents a superposition of two states of opposite momentum, both states having equal amplitudes.

The mean momentum of a particle moving in a potential well according to the laws of classical mechanics is equal to zero: it changes its sign in every reflection from the walls of the well. In this sense we can say that in the case of quantum motion the mean momentum of a particle is also zero. The difference is that at every given instant classical momentum possesses a definite value, whereas the quantum momentum of a particle in a well has no such value: the wave function is represented as a sum of states with momenta of both signs. This corresponds to the uncertainty principle: since the coordinate of the particle is restricted to the limits $0 \leqslant x \leqslant a$, its momentum cannot have an exact value.

Note, furthermore, that in this particular problem of a rectangular well the square of the momentum is equal to $h^2 \varkappa^2$, because the uncertainty extends only to the sign of the momentum. The square of the momentum is in this case proportional to the energy. For a well of arbitrary shape the square of the momentum is also indeterminate.

A Particle in a Three-Dimensional Infinitely Deep Potential Well.

Let us now suppose that a particle is contained in a box whose edges are a_1, a_2, a_3. Generalizing the boundary conditions (28.1), we conclude that the wave function becomes zero on all the sides of

the box:

$$\psi \, (0, \, y, \, z) = \psi \, (x, \, 0, \, z) = \psi \, (x, \, y, \, 0)$$
$$= \psi \, (a_1, \, y, \, z) = \psi \, (x, \, a_2, \, z) = \psi \, (x, \, y, \, a_3) = 0 \qquad (28.13)$$

The wave equation must now be written in three-dimensional form:

$$-\frac{h^2}{2m} \left(\frac{\partial^2 \psi}{\partial x^2} + \frac{\partial^2 \psi}{\partial y^2} + \frac{\partial^2 \psi}{\partial z^2} \right) = E\psi \qquad (28.14)$$

It is convenient to write the solution as follows:

$$\psi = C \sin \varkappa_1 x \sin \varkappa_2 y \sin \varkappa_3 z \qquad (28.15)$$

It is written only in terms of sines and not cosines so as to satisfy the first line of the boundary conditions (28.13). The quantities $\varkappa_1$, $\varkappa_2$, $\varkappa_3$ are determined from the second line of the boundary conditions (28.13). The factors of (28.15) turn zero either at $x = a_1$, or $y = a_2$, or $z = a_3$. In other words

$$\begin{aligned}
\sin \varkappa_1 a_1 &= 0, & \varkappa_1 a_1 &= n_1 \pi \\
\sin \varkappa_2 a_2 &= 0, & \varkappa_2 a_2 &= n_2 \pi \\
\sin \varkappa_3 a_3 &= 0, & \varkappa_3 a_3 &= n_3 \pi
\end{aligned} \qquad (28.16)$$

Here n_1, n_2, and n_3 are integers of which none are equal to zero (otherwise ψ would be equal to zero over all the box).

We substitute (28.15) into (28.14) and take advantage of the fact that an equation of the form

$$\frac{d^2}{dx^2} \sin \varkappa_1 x = -\varkappa_1^2 \sin \varkappa_1 x \qquad (28.17)$$

holds for each term in (25.15), which yields

$$\nabla^2 \psi = -(\varkappa_1^2 + \varkappa_2^2 + \varkappa_3^2) \, \psi$$

For Eq. (28.14) to be satisfied the energy must be related to $\varkappa_1$, $\varkappa_2$, and $\varkappa_3$ in the following way:

$$E = \frac{h^2}{2m} (\varkappa_1^2 + \varkappa_2^2 + \varkappa_3^2) \qquad (28.18)$$

Substituting $\varkappa_1$, $\varkappa_2$, $\varkappa_3$ from (28.16) into (28.18), we obtain the energy eigenvalues

$$E_{n_1 n_2 n_3} = \frac{\pi^2 h^2}{2m} \left(\frac{n_1^2}{a_1^2} + \frac{n_2^2}{a_2^2} + \frac{n_3^2}{a_3^2} \right) \qquad (28.19)$$

The ground state energy is

$$E_{111} = \frac{\pi^2 h^2}{2m} \left(\frac{1}{a_1^2} + \frac{1}{a_2^2} + \frac{1}{a_3^2} \right) \qquad (28.20)$$

The value $E = 0$ is impossible by virtue of the uncertainty principle; a particle contained in a well of finite dimensions has no strictly defined momentum, notably zero momentum. Since $p_x = \pm h \varkappa$, we see that $\Delta p_x = 2h \varkappa$. Substituting Δp_x into the uncertainty relation (22.4), we find that the minimum value of $\varkappa_1$ is π/a_1. But this corresponds precisely to Eq. (28.18) if we substitute $\varkappa_i = \pi/a_i$ $(i = 1, 2, 3)$.

Calculation of the Number of Possible States. To each value of the three numbers n_1, n_2, n_3 there corresponds one possible particle state. Let us find the distribution of the number of possible states of a system over the values of n_1, n_2, and n_3 for large numbers n_1, n_2, n_3.

Numbers that are large in comparison with unity can be differentiated: the differential dn_1 denotes an interval of numbers that is small in comparison with n_1 but still includes many separate integral values of n_1. In other words, $1 \ll dn_1 \ll n_1$. It is then clear that there are exactly dn_1 possible integral numbers in the interval dn_1, and similarly in the intervals dn_2 and dn_3.

Let us plot n_1, n_2, and n_3 on a system of coordinate axes. In this space we construct a parallelepiped with sides dn_1, dn_2, dn_3, so that its volume is equal to $dn_1 dn_2 dn_3$. In accordance with the foregoing, to each point within this parallelepiped, the coordinates n_1, n_2, and n_3 of which are integers, there corresponds one possible particle state in a three-dimensional potential well. There are $dn_1 dn_2 dn_3$ such points within the parallelepiped. Hence, denoting the number of states within the volume $dN(n_1, n_2, n_3)$, we obtain

$$dN(n_1, n_2, n_3) = dn_1 dn_2 dn_3 \qquad (28.21)$$

Substituting $\varkappa_1$, $\varkappa_2$, $\varkappa_3$ from (28.16), we obtain the expression for the number of states in terms of $d\varkappa_1 \, d\varkappa_2 \, d\varkappa_3$:

$$dN(\varkappa_1, \varkappa_2, \varkappa_3) = \frac{a_1 a_2 a_3 \, d\varkappa_1 \, d\varkappa_2 \, d\varkappa_3}{\pi^3}$$

But since $a_1 a_2 a_3$ is the geometrical volume of the box, V, it follows that

$$dN(\varkappa_1, \varkappa_2, \varkappa_3) = \frac{V \, d\varkappa_1 \, d\varkappa_2 \, d\varkappa_3}{\pi^3} \qquad (28.22)$$

The numbers $\varkappa_1$, $\varkappa_2$, $\varkappa_3$ take on only positive values.

In examining the motion of a particle in a one-dimensional potential well we pointed out that to each value of $\varkappa_1$ there correspond two values of the momentum projection, equal in magnitude and opposite in sign. Therefore, if we compare the number of states within the intervals $d\varkappa_1$ and $dp_1/h \equiv dp_x/h$, we find that the latter includes half the number of states of the former. Accordingly, the

number of states in the interval of momentum values $dp_x dp_y dp_z$ is

$$dN(p_x,\ p_y,\ p_z) = \frac{V\,dp_x\,dp_y\,dp_z}{(2\pi h)^3} \tag{28.23}$$

where p_x, p_y, and p_z assume all real values from $-\infty$ to ∞.

In other words, we have managed to discard the rather artificial assumption that a particle must be moving in a box, and the box must be rectangular in shape. Referring, for example, (28.23) to unit volume, we obtain a quite general formula for the corresponding number of states of a quantum particle, the dimension of which is 1/cm³. Note that Eq. (28.23) can be developed only in quantum mechanics, thanks to the finite (nonzero) value of the action quantum.

Equation (28.23) agrees with the uncertainty relation (22.4). If the motion is restricted along x by the interval $(0,\ a_1)$, then only those states differ physically for which the momentum projections differ by not less than $2\pi h/a_1$. Hence, within the interval dp_x there are $dp_x/[2\pi h/a_1] = a_1\,dp_x/(2\pi h)$ states. Finding the product $(a_1 dp_x)/(2\pi h) \times (a_2 dp_y)/(2\pi h) \times (a_3 dp_z)/(2\pi h)$, we arrive at Eq. (28.23). In order to assure that the correct numerical coefficient is obtained in evaluating the number of states from the uncertainty relation the quantity $2\pi h$ was put in the right-hand side of (22.4), or 2π in (19.6a).

Let us now consider the number of states after somewhat changing the variables. We plot the momentum projections p_x, p_y, and p_z on a system of coordinate axes and count the number of states in momentum space lying between two spheres of radii p and $p + dp$. The required number is equal to the integral of (28.23) over the volume contained between the two spheres. The corresponding volume in momentum space is equal to the surface of a sphere of radius p, that is $4\pi p^2$ multiplied by dp:

$$dN(p) \equiv \int dN(p_x,\ p_y,\ p_z) = \frac{V\,4\pi p^2\,dp}{(2\pi h)^3} = \frac{V\,p^2\,dp}{2\pi^2 h^3} \tag{28.24}$$

We now pass from this to the energy of a particle, $E = p^2/(2m)$. Making use of the fact that $p = (2mE)^{1/2}$, $p\,dp = m\,dE$, we find the number of states corresponding to the interval of energy values from E to $E + dE$:

$$dN(E) = \frac{V\,m^{3/2}E^{1/2}\,dE}{2^{1/2}\pi^2 h^3} \tag{28.25}$$

Thus, the number of states in the interval between E and $E + dE$ increases in direct proportion to $E^{1/2}$. In a one-dimensional potential well we would obtain $(am^{1/2}\,dE)/(2^{1/2}\pi h E^{1/2}) \propto E^{-1/2}$.

In courses in mathematical physics it is shown that (28.25) is valid for energy values that are very great in comparison with the energy of the ground state. Then the number of states is proportional

to the volume and does not depend on its shape.

One-Dimensional Potential Well of Finite Depth. We shall now consider the motion of a particle in a one-dimensional potential well of finite depth. We specify it in the following way:

$$U = \infty \quad \text{at} \quad -\infty < x < 0$$
$$= 0 \quad \text{at} \quad 0 \leqslant x \leqslant a$$
$$= U_0 \quad \text{at} \quad a < x < \infty$$

In other words, the potential energy for $x > 0$ is everywhere equal to U_0 except within a region of width a near the coordinate origin,
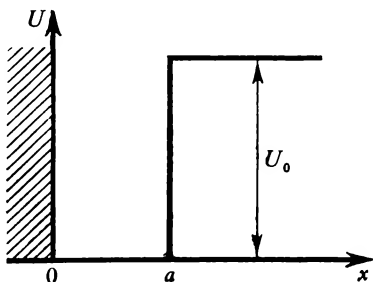
Figure 31

which region we called the well. For $x < 0$ the potential energy is infinite (see Figure 31)[4].

Since the solution is of different analytical form inside and outside the well, we must find the conditions for matching the wave functions at the boundary $x = a$. Let us take the wave equation in the form

$$-\frac{h^2}{2m}\frac{d^2\psi}{dx^2} + U(x)\psi = E\psi \tag{28.26}$$

in which $U(x)$ is defined by the curve in Figure 31, and integrate both sides over a narrow region $a - \delta \leqslant x \leqslant a + \delta$ including the point of discontinuity of the potential energy $x = a$. The integration

---

[4] It was shown in Section 20 how a three-dimensional wave equation can be reduced to a one-dimensional equation by substituting $\psi = \Phi/r$. Then, if the particle's angular momentum is zero, Eq. (24.32) reduces completely to a one-dimensional one, with the difference that now r varies only from zero to infinity. This may be attained formally by situating an infinitely high potential wall at $x = 0$. Figure 31 actually refers to a particle with zero angular momentum in a spherical potential well.

gives

$$-\frac{h^2}{2m}\left[\left(\frac{d\psi}{dx}\right)_{a+\delta}-\left(\frac{d\psi}{dx}\right)_{a-\delta}\right]=\int\limits_{a-\delta}^{a+\delta}(E-U)\,\psi\,dx \qquad (28.27)$$

Even though $U(x)$ suffers a discontinuity at the boundary of the well, it remains finite everywhere. Therefore, when δ approaches zero, the integral on the right also approaches zero. It follows that the left-hand side of (28.27) is also zero. In other words

$$\left(\frac{d\psi}{dx}\right)_{a+0}=\left(\frac{d\psi}{dx}\right)_{a-0} \qquad (28.28)$$

that is, the limit of the derivative on the right is equal to its limit on the left.

This argument would not hold in the problem of an infinitely deep well because then the integral in (28.27) would involve an indeterminate quantity.

We shall now show by a limiting process that the wave function also does not suffer a discontinuity at the boundary. Let us assume the reverse, namely, that ψ suffers a discontinuity Δ: that is $\psi\,(a+0)-\psi\,(a-0)=\Delta$. Furthermore, we assume that the discontinuity occurs not at a point but within some narrow domain close to $x=a$. At points $x-a=\pm\delta$ the wave function ties on smoothly to a certain solution of the wave equation for the narrow transitional domain $a-\delta\leqslant x\leqslant a+\delta$. We thus obtain a stepped curve, only the edges of the steps are rounded. The latter follows from (28.28): the derivative of the wave function cannot have discontinuities at $x=a\pm\delta$.

The order of magnitude of the wave function in the transitional domain is $d\psi/dx\sim\Delta/\delta$, so that in the limit, as $\delta\to0$, it becomes infinite (more precisely, it should become infinite if $\Delta\neq0$!).

Let us now multiply both sides of (28.26) by ψ and transform by parts to get

$$\psi\,\frac{d^2\psi}{dx^2}=\frac{d}{dx}\left(\psi\,\frac{d\psi}{dx}\right)-\left(\frac{d\psi}{dx}\right)^2$$

We integrate this expression between $a-\delta$ and $a+\delta$, obtaining

$$\left(\psi\,\frac{d\psi}{dx}\right)_{a+\delta}-\left(\psi\,\frac{d\psi}{dx}\right)_{a-\delta}-\int\limits_{a-\delta}^{a+\delta}\left(\frac{d\psi}{dx}\right)^2dx$$

$$=-\int\limits_{a-\delta}^{a+\delta}\frac{2m}{h^2}\,(E-U)\,\psi^2\,dx \qquad (28.29)$$

and then perform the limiting process $\delta \to 0$. We may write the integrated terms thus:

$$\left(\psi \frac{d\psi}{dx}\right)_{a+0} - \left(\psi \frac{d\psi}{dx}\right)_{a-0} = [\psi(a+0) - \psi(a-0)]\left(\frac{d\psi}{dx}\right)_{a\pm\delta}$$

because the derivative, as was shown, is not subject to a discontinuity.

Within the assumed discontinuity region of the ψ function, $d\psi/dx$ is of the order of Δ/δ, but at the boundaries of the region it reverts to values independent of δ and therefore finite in accordance with the fact that the edges of the steps are assumed to be rounded. Hence, the whole integrated part on the left in (28.29) is of the order of $(d\psi/dx)_{a\pm\delta}\Delta$. The remaining integral is estimated thus:

$$\int_{a-\delta}^{a+\delta} \left(\frac{d\psi}{dx}\right)^2 dx \sim 2\delta\left(\frac{\Delta}{\delta}\right)^2 \sim \frac{2\Delta^2}{\delta}$$

Hence it tends to infinity as δ tends to zero. The right-hand side of (28.29) is finite for $\delta \to 0$. Thus, the assumption that $\psi(x)$ has a finite discontinuity Δ results in an infinite term in Eq. (28.29), which cannot be cancelled out, that is, in a contradiction. To eliminate the contradiction we must assume that $\Delta = 0$, in other words, the wave function does not suffer a discontinuity.

Thus, at the discontinuities of the potential energy curve the wave function is continuous together with its first derivative.

Actually, we know of no interactions in nature that would correspond to potential energy curves with discontinuities. However, there are forces that decrease very rapidly with distance, specifically, nuclear forces. As yet we are unable to formulate an exact law of their dependence upon distance, but situations are well known when these forces pass from very large values to zero over distances substantially smaller than the de Broglie wavelength of nuclear particles. In such cases the dependence of the force upon distance can be legitimately approximated by the stepped curve in Figure 31. Such an approximation yields quite reasonable results in a number of cases.

But having assumed the existence of a discontinuity in the potential energy curve, it was necessary to investigate the behaviour of the wave function in the neighbourhood of the discontinuity, so that the model of a discontinuous potential curve would be intrinsically consistent and not lead to mathematical contradictions.

Having the boundary conditions, we can now develop a solution for the wave equation. The wave equation for the region $0 \leqslant x \leqslant a$ (inside the well) is of the form

$$-\frac{h^2}{2m} \frac{d^2\psi}{dx^2} = E\psi$$

We take its solution

$$\psi = C_1 \sin \varkappa x \qquad (28.30)$$

where $\varkappa$ is defined from (28.3). The solution involving the sine only is taken because at the left edge of the well, where the potential energy suffers an infinite discontinuity, ψ satisfies the boundary condition (28.1), $\psi(0) = 0$.

The wave equation outside the well, where $x > a$, is

$$-\frac{h^2}{2m}\frac{d^2\psi}{dx^2} = (E - U_0)\psi \qquad (28.31)$$

(i) First we consider the case $E > U_0$. Introducing the abbreviated notation

$$\frac{2m}{h^2}(E - U_0) \equiv \varkappa_1^2 \qquad (28.32)$$

we obtain (28.31) in the standard form (28.4):

$$\frac{d^2\psi}{dx^2} = -\varkappa_1^2 \psi$$

whence

$$\psi = C_2 \sin \varkappa_1 x + C_3 \cos \varkappa_1 x \qquad (28.33)$$

We must now satisfy the boundary conditions on the right edge of the potential well, where $U(x)$ suffers only a finite discontinuity. According to these conditions both the wave function and its first derivative are continuous:

$$C_1 \sin \varkappa a = C_2 \sin \varkappa_1 a + C_3 \cos \varkappa_1 a \qquad (28.34)$$

$$\varkappa C_1 \cos \varkappa a = \varkappa_1 C_2 \cos \varkappa_1 a - \varkappa_1 C_3 \sin \varkappa_1 a \qquad (28.35)$$

From these equations we can determine C_2 and C_3 in terms of C_1, thereby matching the solution of the wave equation outside the well with the solution inside the well. Equations (28.34) and (28.35) are linear with respect to C_2 and C_3 and have solutions for all values of the coefficients:

$$C_2 = \frac{\varkappa_1 \sin \varkappa a \sin \varkappa_1 a + \varkappa \cos \varkappa a \cos \varkappa_1 a}{\varkappa_1} C_1$$

$$C_3 = \frac{\varkappa_1 \sin \varkappa a \cos \varkappa_1 a - \varkappa \cos \varkappa a \sin \varkappa_1 a}{\varkappa_1} C_1$$

The only exception is the value of E at which $\varkappa_1 = 0$, that is $E = U_0$. This point does not belong to the spectrum of the permitted energy values. But at $E > U_0$ the Schrödinger equation always has a solution. There is no discreteness in the eigenvalue spectrum.

We can choose the gauge of the potential energy in this problem so that it is zero at $x = \infty$, that is, consider it equal to zero for

$x > a$ and equal to $-U_0$ for $0 \leqslant x \leqslant a$. Then the case which we have just considered corresponds to positive eigenvalues of the total energy.

(ii) Now let $E < U_0$. We introduce the quantity

$$\frac{2m}{h^2}(U_0 - E) \equiv \kappa^2 \qquad (28.36)$$

The wave equation is now written differently than for $E > U_0$, namely

$$\frac{d^2\psi}{dx^2} = \kappa^2 \psi$$

and its solution is expressed in terms of the exponential function

$$\psi = C_4 e^{\kappa x} + C_5 e^{-\kappa x} \qquad (28.37)$$

But the exponential $e^{\kappa x}$ tends to infinity as x increases. For $x = \infty$ it would give an infinite probability for finding the particle, and no finite value could be assigned to the integral $\int\limits_0^\infty |\psi|^2 \, dx$. It follows that a physically meaningful solution exists only for $C_4 = 0$ and must be of the form

$$\psi = C_5 e^{-\kappa x} \qquad (28.38)$$

Let us again try to satisfy the boundary conditions at $x = a$. This time they appear as follows:

$$C_1 \sin \kappa a = C_5 e^{-\kappa a} \qquad (28.39)$$

$$\kappa C_1 \cos \kappa a = -\kappa C_5 e^{-\kappa a} \qquad (28.40)$$

We divide equation (28.40) by (28.39) in order to eliminate C_1 and C_5, and obtain

$$\kappa \cot \kappa a = -\kappa \qquad (28.41)$$

From this equation we find the expression for $\sin \kappa a$:

$$\sin \kappa a = \pm (1 + \cot^2 \kappa a)^{-1/2} = \pm \left[1 + \left(\frac{\kappa}{\varkappa} \right)^2 \right]^{-1/2}$$

$$= \pm \left[1 + \frac{U_0 - E}{E} \right]^{-1/2} = \pm \left(\frac{E}{U_0} \right)^{1/2} \qquad (28.42)$$

Let us reduce this equation to a more convenient form. From (28.3)

$$E^{1/2} = \frac{h}{a \, (2m)^{1/2}} \varkappa a$$

so that

$$\sin \varkappa a = \pm \frac{h}{a \, (2mU_0)^{1/2}} \varkappa a \qquad (28.43)$$

Only those solutions should be chosen for which cot $\varkappa a$ is negative, in accordance with (28.41). Hence, $\varkappa a$ must lie in the second, fourth, sixth, eighth, and in general only even, quadrants.

We shall solve Eq. (28.43) graphically (Figure 32). The left-hand side of the equation is represented by a sinusoid, while the right-hand side is represented by two straight lines with the slopes $\pm h\,[a\,(2mU_0)^{1/2}]^{-1}$. If the absolute value of the tangent of the angle of inclination of these lines is less than $2/\pi$, they have one or
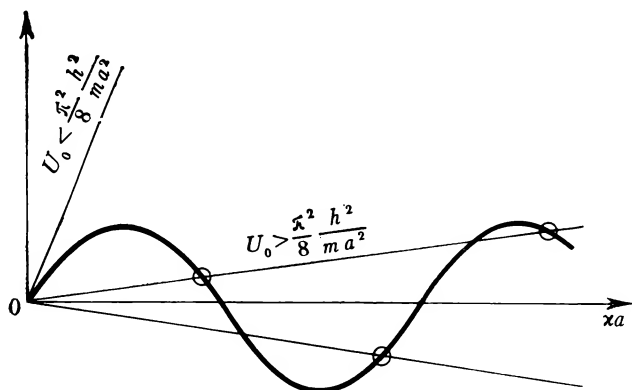
Figure 32

several common points with the sinusoid in the quadrants corresponding to the roots of (28.41). The trivial point of intersection, $\varkappa a = 0$, does not count because at $\varkappa = 0$ the wave function vanishes everywhere. Thus, in a well of finite depth of the form considered there are only several energy eigenvalues. But if

$$U_0 < \frac{\pi^2}{8}\,\frac{h^2}{ma^2} \equiv U_{\mathrm{cr}} \qquad\qquad (28.44)$$

there are in general no points of intersection of the straight lines with the sinusoid corresponding to energy eigenvalues (the intersection point in the first quadrant does not count!). In Figure 32 the points of intersection in the even quadrants are marked by small circles.

Of special interest is the case when the energy level lies very close to the edge of the potential well in comparison with its total depth. Suppose that this is the ground level and there are no other levels in the well. Then, assuming the width of the well to be given and the depth to be slightly greater than the critical value, U_{cr}, we write

$$U_0 = U_{\mathrm{cr}}\,(1 + v), \qquad \varkappa a = \frac{\pi}{2}\,(1 + y)$$

where v and y are small quantities. Since close to a maximum value the sine differs from unity by a second-order quantity, we obtain the relationship between v and y:

$$\sin \varkappa a \approx 1 = \frac{1+y}{(1+v)^{1/2}}, \qquad y \approx \frac{v}{2}$$

Now, making use of the definition of $\varkappa$, (28.3), we represent Eq. (28.43) as

$$\sin \varkappa a = \sin(1+y)\frac{\pi}{2} \approx 1 - \frac{\pi^2}{8}y^2 = \left(\frac{E}{U_0}\right)^{1/2} \equiv \left(\frac{U_0-\varepsilon}{U_0}\right)^{1/2}$$

Assuming E to differ from U_0 by the small quantity ε, we find that

$$\frac{\varepsilon}{U_0} = \frac{\pi^2}{4}y^2 = \frac{\pi^2}{16}v^2$$

whence it is apparent that

$$\frac{\varepsilon}{U_0} = \frac{\pi^2}{16}\left(\frac{U_0}{U_{\mathrm{cr}}} - 1\right)^2 \tag{28.45}$$

Hence, the level corresponds to the condition $\varepsilon \ll U_0$: if the depth of the well differs from the limiting depth at which the level just appeared by a first-order quantity, the energy level lies at a distance of the second order from the upper edge of the well.

Such a case is actually obtained in a heavy hydrogen nucleus, the deuteron. The depth of the potential well corresponding to the nuclear forces is estimated as 20-30 MeV, while the proton-neutron binding energy, ε, is about 2.2 MeV (we recall that the two-body problem (the proton plus a neutron) reduces to a one-body problem (see Sec. 3)). Thanks to the small binding energy of the particles in a deuteron, the wave function of their relative motion outside the well has, according to Eq. (28.38), as well as (28.44) and (28.45), the following form:

$$\psi \approx C_5 \exp\left[-\frac{x}{a}\left(\frac{U_0}{U_{\mathrm{cr}}} - 1\right)\frac{\pi^2}{8}\right] \tag{28.46}$$

But it is apparent from this equation that the function ψ falls off rapidly at a distance of order

$$\frac{8a}{\pi^2}\left(\frac{U_0}{U_{\mathrm{cr}}} - 1\right)^{-1}$$

which is many times greater than the dimensions of the potential well (to be more precise, in the case of a deuteron it is not many times, but two or three times). But then the integral taken over the domain outside the well, corresponding to $(U_0/U_{\mathrm{cr}} - 1)^{-3}$, exceeds the integral over the domain within the well. In other words, the probability of finding a particle outside the domain of nuclear forces is much greater than of finding it within that domain. That is

23*

why the wave function is known much better than the proton-neutron interaction. But the properties of a deuteron are calculated from its wave function, not directly from its nuclear forces. This was used by H. Bethe and R. Peierls to develop a quite satisfactory quantitative theory of the deuteron.

Finite and Infinite Motion. We shall now show that the shape of the energy spectrum is related to the type of motion. For $E > U_0$ the solution outside the well is of the form (28.33). It remains finite also for infinitely large x. Therefore, the integral $\int |\psi|^2 \, dx$ taken over the region of the well is infinitesimal compared with the same integral taken over all the space. In other words, there is nothing to prevent the particle from going to infinity. Such motion was termed infinite in Section 5.

For $E < U_0$, solution (23.38), when it exists, falls off exponentially as $x \to \infty$. Hence, the probability of the particle receding to an infinite distance from the origin is zero: the particle all the time remains at a finite distance from the well. It is natural to call such motion finite, as in classical mechanics.

Thus, infinite motion has a continuous energy spectrum, while finite motion has a discrete spectrum comprising separate values. If the depth of the well is very small, there may be no finite motion. The latter holds only in the three-dimensional case: in a one-dimensional or two-dimensional well there is always at least one bound level with negative total energy. That is why we emphasized that the problem of a potential well of finite depth in effect refers to three-dimensional motion. In classical mechanics, at $E < U_0$ motion is in every case finite in a well of any number of dimensions, including a three-dimensional well.

In the course of the solution it becomes apparent that the obtained result refers not only to a rectangular potential well. Indeed, if the potential energy is gauged to zero at infinity, then the solution with positive total energy is of the form (28.33) for sufficiently large x, while the solution with negative total energy is of the form (28.38). The latter contains only one arbitrary constant, while (28.33) contains two constants. The integral curves of both solutions must be extended to the coordinate origin in order that the condition $\psi(0) = 0$ can be satisfied (we consider that x is always greater than zero). Obviously, if we have two constants at our disposal, we can always choose them so that the condition $\psi(0) = 0$ is satisfied[5]. But a solution of the form (28.38) containing one constant becomes zero at the origin only for certain special values of $\varkappa$.

---

[5] If $\psi(0) = C_1\psi_1(0) + C_2\psi_2(0)$, then $C_1/C_2 = -\psi_2(0)/\psi_1(0)$.

We can also offer another explanation of why infinite motion has a continuous energy spectrum. The wave function of a particle in infinite motion differs from the wave function of a free particle only in the region of a potential well. But the probability of finding the particle in this region is infinitesimal if the whole region of motion is sufficiently large. Therefore, the wave function for infinite motion coincides with the wave function of a free particle in "almost" the entire space, for which the probability of finding the particle is equal to unity, and the energy spectrum turns out to be the same as for a free particle.

If U_0 tends to infinity, the wave function outside the well falls off very rapidly. In the limit ($U_0 \to \infty$) it vanishes at an infinitesimal distance from the boundary $x = a$, yielding the boundary condition (28.1).

In the case of finite U_0 the wave function outside the well does not become zero at once. Therefore, a nonzero probability exists that the particle will be outside the well at a finite distance from it.

This would have been quite impossible in classical mechanics, as is obtained from (28.38) in the limiting transition $h \to 0$. In this case $k = \infty$, and ψ becomes infinitesimal outside the well. This, naturally, should be the case: if the particle is situated outside the well, its kinetic energy (in the classical sense) is $E - U_0 < 0$. But the velocity of such a particle is an imaginary quantity. In classical mechanics this means that a given point of space is absolutely unattainable for a particle with energy E.

In quantum mechanics, position and velocity never exist in the same states as precise quantities. Earlier we interpreted this in terms of the uncertainty relation, that is, we considered cases for which precision in the concept of velocity for a certain state was restricted by the limits $2\pi h/(m\Delta x)$. But this is a lower limit and has to do with particles which are almost unaffected by forces. The appearance of an imaginary velocity in the equation for a bound particle shows that the very concept of velocity is not applicable to a region of space, however large, for which $U > E$. We can express this differently by saying that, for $U > E$, the uncertainty in the kinetic energy is always greater than the difference $U - E$.

We have seen on the example of the deuteron that a bound particle may, even with overwhelming probability, occur in a domain where no forces are acting.

Thus, in classical mechanics there is no analogue for the motion of a bound particle outside a well. And that is precisely the domain of motion that is decisive in finding the energy spectrum for finite motion.

The Linear Harmonic Oscillator. We shall consider the problem of a quantum linear harmonic oscillator. We already know its

Hamiltonian from Eq. (27.16). After replacing the momentum opera-
tor by $(h/i)(d/dx)$, we obtain the Schrödinger equation

$$-\frac{h^2}{2m}\frac{d^2\psi}{dx^2}+\frac{m\omega^2 x^2}{2}\,\psi=E\psi \tag{28.47}$$

Let us now introduce other units of measurement, in particular,
we shall take the unit of length equal to $[h/(m\omega)]^{1/2}$, so that

$$x=\left(\frac{h}{m\omega}\right)^{1/2}\xi \tag{28.48}$$

The quantity ξ is dimensionless. The derivative $d\psi/dx$ is equal to

$$\frac{d\psi}{dx}=\left(\frac{m\omega}{h}\right)^{1/2}\frac{d\psi}{d\xi}$$

Further, we put

$$2E=\varepsilon h\omega$$

In terms of these dimensionless variables, Eq. (28.47) assumes
the form

$$-\frac{d^2\psi}{d\xi^2}+\xi^2\psi=\varepsilon\psi \tag{28.49}$$

Equation (28.49) does not contain any parameters of the problem,
that is ω, m, or h. For this reason the eigenvalue ε can only be an
abstract number. Comparing this with the expression for energy,
we see that the energy eigenvalue of a harmonic oscillator is pro-
portional to its frequency ω.

To solve Eq. (28.49), it is convenient to introduce a new dependent
variable $g(\xi)$ such that

$$\psi=g(\xi)\,e^{-\xi^2/2}$$

whence

$$\frac{d\psi}{d\xi}=-\xi g(\xi)\,e^{-\xi^2/2}+\frac{dg}{d\xi}\,e^{-\xi^2/2}=e^{-\xi^2/2}\left(-\xi g(\xi)+\frac{dg}{d\xi}\right)$$

$$\frac{d^2\psi}{d\xi^2}=e^{-\xi^2/2}\left(\xi^2 g(\xi)-g(\xi)-2\xi\frac{dg}{d\xi}+\frac{d^2g}{d\xi^2}\right)$$

Substituting these relations into (28.49) and carrying out the neces-
sary cancellations, we obtain the equation for the new dependent
variable:

$$-\frac{d^2g(\xi)}{d\xi^2}+2\xi\frac{dg(\xi)}{d\xi}=(\varepsilon-1)\,g(\xi) \tag{28.50}$$

The coefficients in this equation contain ξ in not higher than the
first power and it is therefore comparatively simple to integrate it.

We seek the solution in the form of the power series

$$g(\xi) = g_0 + g_1\xi + g_2\xi^2 + g_3\xi^3 + \ldots = \sum_{n=0}^{\infty} g_n\xi^n$$

In order to determine the expansion coefficients g_n, we must substitute the series into Eq. (28.49), differentiate it by terms and compare the expressions for the same powers of ξ^n. The first derivative is

$$\frac{dg(\xi)}{d\xi} = g_1 + 2g_2\xi + 3g_3\xi^2 + \ldots = \sum_{n=1}^{\infty} ng_n\xi^{n-1}$$

so that

$$2\xi\frac{dg(\xi)}{d\xi} = \sum_{n=0}^{\infty} 2ng_n\xi^n$$

The second derivative is

$$\frac{d^2g(\xi)}{d\xi^2} = 2g_2 + 6g_3\xi + \ldots = \sum_{k=2}^{\infty} (k-1)kg_k\xi^{k-2}$$

In the last summation we changed the summation index, denoting it by the letter k. We shall now revert to n, assuming that $k - 2 = n$, or $k = n + 2$. Then

$$\frac{d^2g(\xi)}{d\xi^2} = \sum_{n=0}^{\infty} (n+2)(n+1)g_{n+2}\xi^n$$

Now substituting the expressions for the first and second derivatives into Eq. (28.50) and collecting coefficients of ξ^n, we obtain

$$\sum_{n=0}^{\infty} \xi^n\left[-(n+2)(n+1)g_{n+2} + 2ng_n - (\varepsilon-1)g_n\right] = 0$$

We know that for a power series to be equal to zero, the coefficients of ξ^n must vanish. Thus

$$g_{n+2} = g_n\frac{2n+1-\varepsilon}{(n+2)(n+1)} \tag{28.51}$$

In this way the expansion proceeds in powers of ξ^2 because the coefficients g_n go alternately.

Let us assume initially that $g_0 \neq 0$. Then, from Eq. (28.51), we find in turn $g_2, g_4, \ldots, g_{2k}$. Not a single odd coefficient appears in the series if $g_1 = 0$. From the recurrence formula (28.51), all g_{2k+1} successively vanish. Conversely, if $g_0 = 0$, $g_1 \neq 0$, then only the coefficients with odd indices remain in the series. It is therefore

sufficient to examine the solutions containing either only even or only odd powers of ξ^2. To be specific, let us take the series of even powers.

Let us examine the behaviour of the series for large values of ξ. Then terms involving high powers of ξ, that is, large n, are predominant. But if n is a large number, then we can neglect the constant numbers in comparison with n in the terms appearing in the recurrence formula (28.51), obtaining the asymptotic relation

$$g_{n+2} = \frac{2}{n} g_n$$

Since we have agreed to assume n to be an even number, we put $n' = 2n$. Instead of g_{2n} we write $g_{2n} \equiv g_{n'}$, so that $g_{n'}$ is the coefficient of the series in powers of ξ^2. For these coefficients the asymptotic recurrence relation is written as follows:

$$g_{n'+1} = \frac{g_{n'}}{n'}$$

In the function with odd powers of ξ, we can take ξ outside the parentheses and then repeat the derivation of the asymptotic relation between the coefficients of the series in the parentheses. Obviously, we will obtain the same relationship as for even n.

It is now easy to find the form of the coefficients themselves for large n:

$$g_{n'+1} = \frac{g_0}{n(n-1)(n-2)\ldots 1} = \frac{g_0}{n!}$$

although actually we have in the denominator not exactly $n!$, because at n' close to unity the very recurrence relation between the coefficients is not correct. The way $g_{n'}$ decreases when n' increases is correct for large n'. Using the obtained expression for $g_{n'}$, we find the asymptotic expansion for the function $g(\xi)$:

$$g(\xi) = \sum_{n'=0}^{\infty} g_{n'} (\xi^2)^{n'} = \sum_{n'=0}^{\infty} \frac{(\xi^2)^{n'}}{n'!} = e^{\xi^2}$$

Thus, the asymptotic behaviour of $g(\xi)$ is described by the exponential function e^{ξ^2}. But then it is also possible to find the wave function ψ for large ξ:

$$\psi(\xi) = e^{\xi^2 - \xi^2/2} = e^{\xi^2/2}$$

However, this form of $\psi(\xi)$ is quite unacceptable: ψ must remain finite at large ξ and not increase.

There is only one possibility of obtaining a finite value of $\psi(\xi)$ at infinity. For that the series for $g(\xi)$ must terminate at some n, and all subsequent coefficients g_{n+2}, g_{n+4}, etc. should be identically

equal to zero. From Eq. (28.51) it is apparent that g_{n+2} becomes zero when

$$\varepsilon = 2n + 1$$

where n is any integer or zero. Since g_{n+4} is linearly expressed in terms of g_{n+2}, it is sufficient for g_{n+2} to vanish for the series to terminate at g_n. It follows that at $\varepsilon = 2n + 1$ the function $g(\xi)$ becomes a polynomial. The product of the polynomial multiplied by the exponential $e^{-\xi^2/2}$ always tends to zero as $\xi \to \infty$. Hence $\psi(\infty) = 0$.

Here, the wave function of an oscillator corresponds to its finite motion in the same sense as in classical mechanics: the probability that a particle will recede to infinity is zero. A discrete energy spectrum corresponds to finite motion. From the expression for ε we find

$$E_n = h\omega \left(n + \frac{1}{2} \right) \tag{28.52}$$

The least possible energy is $E_0 = h\omega/2$. All this agrees with the results obtained in the preceding section.

The state with energy E_0, as mentioned before, is called the ground state. The function $g(\xi)$ of this state terminates already at the zero term, so that the corresponding wave function has the form

$$\psi = g_0 e^{-\xi^2/2} \tag{28.53}$$

This function does not have any zeros at a finite distance from the coordinate origin, which must be the case for the ground state.

It may be noted that the state with zero energy would correspond to a particle at rest at the origin. However, such a state is not compatible with the uncertainty principle, since an oscillator in it would have simultaneously a coordinate and velocity.

Let us also find the eigenfunctions of the first and second excited states. In the first state $E_1 = h\omega (1 + 1/2) = 3h\omega/2$. Here we must put $g_0 = 0$, since the first term of the series $g(\xi) = g_1\xi$ is not zero, $\varepsilon = 3$, and all the other coefficients g_{2n+1} with odd numbers vanish. The even coefficients are absent altogether, because $g_0 = 0$. In general, all functions with even n turn out to be even, $\psi(-\xi) = \psi(\xi)$, and with odd n they are odd, $\psi(-\xi) = -\psi(\xi)$. As was just shown, the function with $n = 1$ has the form

$$\psi_1 = g_1 \xi e^{-\xi^2/2}$$

It becomes zero precisely at $\xi = 0$, that is, has one node.

In the same way it is easy to find the function ψ_2. Indeed, $E_2 = h\omega (2 + 1/2) = 5h\omega/2$, $\varepsilon = 5$. From the exact recurrence formula, the coefficient g_2 is

$$g_2 = g_0 \frac{1-\varepsilon}{1\cdot 2} = -2g_0$$

whence
$$\psi_2 = g_0 (1 - 2\xi^2) e^{-\xi^2/2}$$

The nodes of this function are situated at the points $\xi = \pm 1/\sqrt{2}$. In general, the function ψ_n has n nodes. The functions for the first few n are shown in Figure 33.
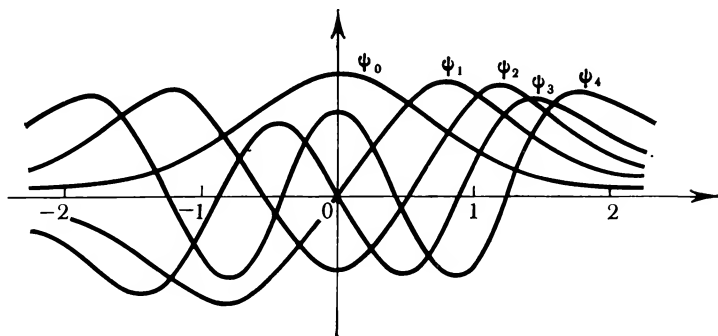
Figure 33

The energy eigenvalue distribution of a linear harmonic oscillator and its potential energy curve are shown in Figure 34. It is very interesting that energy eigenvalues are separated by equal intervals.
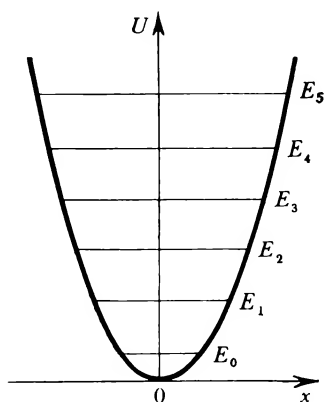
Figure 34

The oscillator problem qualitatively resembles the problem of a rectangular well of infinite depth, but in a well the energy of a level increases in proportion to the square of the number of the respective level.

More like the problem of determining the energy levels in an infinitely deep potential well is the problem of determining the eigenvalues of the angular momentum square. In both cases the operator whose eigenvalues are being determined is expressed in terms of the second derivatives, and the independent variables vary within a restricted (finite) domain, $0 \leqslant x \leqslant a$ for the well, and $0 \leqslant \vartheta \leqslant \pi$, $0 \leqslant \varphi \leqslant 2\pi$ for the angular momentum square. Therefore, in both cases the eigenvalues depend on the number quadratically:

$$E_n = \frac{\pi^2 h^2 n^2}{2ma^2}, \qquad M_l^2 = h^2 l\,(l+1)$$

EXERCISES

1. A potential-energy curve is given as follows: the potential energy is zero for $x < 0$, and is equal to U_0 for $x \geqslant 0$ (the potential step). A beam of particles is directed from the left along x, that is, the motion is one-dimensional. Determine the reflection factor at $E > U_0$.

Solution. The wave function on the left from the step is

$$C_1 e^{ipx/h} + C_2 e^{-ipx/h}$$

On the right (above the step) the function must be sought in the form

$$C_3 e^{ip'x/h}, \qquad \text{where} \quad p' = [2m\,(E-U_0)]^{1/2}$$

since by definition in this domain there are no particles moving in the direction of negative x. Note that the wave $e^{-i(Et-px)/h}$ travels from left to right, and the wave $e^{-i(Et+px)/h}$ from right to left (see Sec. 18).

From the boundary conditions at $x = 0$ we find the ratio $|C_2/C_1|^2$, that is, the ratio of the squares of the amplitudes of the reflected and incident waves, which is in fact the reflection factor

$$\left| \frac{C_2}{C_1} \right|^2 = \left(\frac{p-p'}{p+p'} \right)^2$$

For $E < U_0$ we see that p' is a purely imaginary quantity, and the reflection factor is always unity.

2. A potential-energy curve is given as follows: the potential energy is zero for $x < 0$ and for $x > a$. For $0 \leqslant x \leqslant a$, it is equal to U_0 such that $U_0 > 0$ (the potential barrier). A beam of particles whose energy is less than U_0 impinges from the left. Determine the reflection factor.

Solution. The wave function to the left of the barrier is equal to $e^{ikx} + Ce^{-ikx}$ ($k = p/h$). For simplicity, we put $C_1 = 1$, since we are interested only in the ratio $C_2/C_1 = C$. Below the barrier, the wave function has the form of a sum of the exponents of a real argument, $C_1' e^{\varkappa x} + C_2' e^{-\varkappa x}$; beyond the barrier we again seek the wave function in the form of a wave travelling from left to right, $C_3 e^{ikx}$ (there we have only the transmitted wave, while

before the barrier we have the incident and reflected waves). The constants C, C_1', C_2', and C_3 are determined from the continuity conditions for the wave function and its derivative at the boundaries of the barrier. The expressions for the constants C and C_3 have, accordingly, the form

$$C = \frac{2(\varkappa^2 + k^2)\sinh \varkappa a}{(\varkappa + ik)^2 e^{-\varkappa a} - (\varkappa - ik)^2 e^{\varkappa a}}$$

$$C_3 = \frac{4ik\varkappa e^{-ika}}{(\varkappa + ik)^2 e^{-\varkappa a} - (\varkappa - ik)^2 e^{\varkappa a}}$$

and the expressions for the flux on both sides of the barrier are

$$j = \frac{hk}{m}(1 - |C|^2), \qquad j = \frac{hk}{m}|C_3|^2$$

Substituting $|C|$ and $|C_3|$, we find that both expressions for the flux coincide, as could have been expected.

If $\varkappa a \gg 1$, that is, the barrier is almost opaque, we have

$$C \approx -1, \qquad C_3 \approx -\frac{4ik}{\varkappa} e^{-ika} e^{-\varkappa a}$$

Thus the transmitted flux decreases exponentially with the thickness of the barrier. Note also that if $E > U_0$, that is, the energy of the particles lies above the barrier, some will nevertheless be reflected ($|C_3| < 1$). In classical mechanics such a barrier does not reflect.

3. Show that reflection occurs from a potential well for which $U = 0$ at $-\infty < x < 0$, $U = -|U_0|$ at $0 \leqslant x \leqslant a$, and $U = 0$ at $a < x < \infty$.

4. Verify the orthogonality property of the wave functions for a well of infinite and finite depth.

5. Show that the functions $g_n(\xi)$, in terms of which the wave functions of a linear harmonic oscillator are expressed, can be expressed to the accuracy of the constant factor in the form

$$g_n(\xi) = e^{\xi^2}\left(\frac{d}{d\xi}\right)^n e^{-\xi^2}$$

Verify this by substitution into Eq. (28.50) at $\varepsilon = 2n + 1$.

6. Normalize the functions ψ_0 and ψ_1 of the harmonic oscillator, taking advantage of the fact that

$$\int_{-\infty}^{\infty} e^{-\xi^2} d\xi = \sqrt{\pi}, \qquad \int_{-\infty}^{\infty} \xi^2 e^{-\xi^2} d\xi = \sqrt{\pi}/2$$

7. Verify that in the ground state of a linear harmonic oscillator $\langle(\Delta p)^2\rangle\langle(\Delta x)^2\rangle$ assumes the maximum value $h^2/4$ (see (25.21)).

29

MOTION IN A CENTRAL POTENTIAL

The motion of an electron in a central attractive field is the principal problem in the quantum mechanics of the atom. And it is not necessary to regard the field as strictly Coulomb in character. For example, in alkali-metal atoms, an outer electron which is bound relatively weakly to the nucleus moves in the field of the nucleus and the so-called atomic core, that is, all the other electrons. Such an approximate description provides a satisfactory understanding of the peculiarities of the behaviour of alkali-metal atoms and their energy states without solving the extremely difficult many-body problem of quantum mechanics. Even in those cases when the replacement of an exact description by the approximate concept of the resultant field of the rest of the electrons acting on the given electron is unsatisfactory from the quantitative aspect, the general qualitative picture of the atomic state is nevertheless conveyed correctly and helps in the classification of separate states.

The approximate approach is successful because it correctly represents the spatial distribution of the nodal surfaces of the wave function. Therein lies the significance of the general problem on the motion of an electron in a central field. Such a field is described by a certain function (potential) $U(r)$ whose form need not be specified in the most general case. However, in the immediate vicinity of the force centre (the nucleus), where the screening effect of the other electrons on the given one is least, $U(r)$ tends to infinity according to the Coulomb law. Furthermore, we assume that $U(\infty) = 0$.

The Eigenfunctions of the Angular Momentum Square and Projection (Spherical Functions). Referring to (24.35), we write the Hamiltonian of a particle moving in a central force field as follows:

$$\hat{\mathscr{H}} = -\frac{h^2}{2m}\frac{1}{r^2}\frac{\partial}{\partial r}r^2\frac{\partial}{\partial r} + \frac{\hat{M}^2}{2mr^2} + U(r) \tag{29.1}$$

Since $\hat{M}^2$ involves differentiations only with respect to angles, it commutes with the energy operator $\hat{\mathscr{H}}$. Hence, in one and the same state we have eigenvalues for the energy, the angular momentum square M^2, and the angular momentum projection M_z, which commutes with both the Hamiltonian and the angular momentum square.

However, $\hat{M}_z$ is not involved in the Hamiltonian, hence M_z may have different eigenvalues hk for one and the same energy E of the system. To the same energy E there corresponds a whole set of eigen-

functions ψ_{Elk}, where $-l \leqslant k \leqslant l$. States described by such wave functions are termed *degenerate*.

It is not difficult to establish the criterion of *degeneracy*, that is, the conditions at which it appears. It was shown in Section 24 that different angular momentum components, for example M_x and M_z, do not commute, but they commute with the angular momentum square and, as can be seen from (29.1), with the Hamiltonian of motion in a central field. Let us operate with $\hat{M}_x$ on the function ψ_{Elk}. Since it is not an eigenfunction of the operator, we get an expansion involving all the eigenfunctions of $\hat{M}_z$:

$$\hat{M}_x \psi_{Elk} = \sum_{k'=-l}^{l} c_{kk'} \psi_{Elk'} \tag{29.2}$$

(As was shown in Section 25, the square of the expansion coefficient modulus, $|c_{k,\,k'}|^2$, is equal to the probability that, in passing through a magnetic field directed along the x axis, a particle with the given angular momentum projection $M_z = hk$ will be found in a beam where $M_x = hk'$.) The wave function (29.2), which is equal to $\hat{M}_x \psi_{Elk}$, is at the same time the eigenfunction of the Hamiltonian (29.1) because $\hat{M}_x$ commutes with the Hamiltonian; each term $\psi_{Elk'}$ is an eigenfunction of $\hat{\mathscr{H}}$, but not of $\hat{M}_z$. Thus, because the operators $\hat{M}_x$ and $\hat{M}_z$ do not commute, different wave functions, ψ_{Elk} and $\hat{M}_x \psi_{Elk}$, correspond to the same energy, that is, degeneracy occurs.

It is obvious that this reasoning is applicable to the eigenfunctions of any operator $\hat{\lambda}$, provided there are two other operators, $\hat{\mu}$ and $\hat{\nu}$, that commute with $\hat{\lambda}$ and do not commute between themselves: $\hat{\mu}\psi_{\lambda\nu}$ does not coincide with $\psi_{\lambda\nu}$, but at the same time is an eigenfunction of $\hat{\lambda}$, since $\hat{\mu}\hat{\lambda}\psi = \hat{\lambda}(\hat{\mu}\psi) = \lambda(\hat{\mu}\psi)$.

In Section 24 we developed the eigenfunctions of the operators $\hat{M}^2$ and $\hat{M}_z$ and represented them in the form (24.39). Here we shall express the same functions in a somewhat different manner more convenient for direct computations.

Restricting ourselves to the case of $k = 0$, we find that the wave function (24.39) can, with the addition of a numerical factor, be represented as

$$Y_l^0 = \frac{(-1)^l}{l!} r^{(l+1)} \left(\frac{\partial}{\partial z}\right)^l \frac{1}{r} \tag{29.3}$$

The factor was introduced for the definition of Y_l^0 to coincide with standard notation. It can be seen from the form of (29.3) that this formula is an expression of a homogeneous coordinate function of zero dimensions. Hence, it is a function only of angles, as it should be. Furthermore, since the wave function (29.3) involves only

differentiation with respect to z, it contains only the angle between z and r, or more precisely, its cosine ($\cos \vartheta$). Thus, the function defined by (29.3) is a polynomial of $\cos \vartheta$.

This polynomial satisfies the following differential equation:

$$\frac{1}{\sin \vartheta} \frac{d}{d\vartheta} \sin \vartheta \frac{d}{d\vartheta} Y_l^0$$

$$= \frac{d}{d\cos \vartheta} (1 - \cos^2 \vartheta) \frac{d}{d\vartheta} Y_l^0 = -l(l+1) Y_l^0 \quad (29.4)$$

which is obtained from (24.35) if we put $p_\varphi = M_z = hk = 0$, since the eigenfunction of the angular momentum square, $Y_l^0 (\cos \vartheta)$, corresponds to the zero value of the angular momentum projection on the z axis.

Equation (29.4) is a second-order equation, hence it has two linearly independent solutions. However, only one of them, namely $Y_l^0 (\cos \vartheta)$, is regular, that is, has no singular points. It is not hard to verify that in the general case (for arbitrary l) the second solution has singular points at $\cos \vartheta = \pm 1$, but we shall restrict ourselves to the proof for $l = 0$. The regular solution is $Y_0^0 = 1$, while the second solution, as can be seen from (29.4), is determined as follows:

$$\frac{d}{d\cos \vartheta} (1 - \cos^2 \vartheta) \frac{d\widetilde{Y}_0^0}{d\cos \vartheta} = 0, \qquad (1 - \cos^2 \vartheta) \frac{d\widetilde{Y}_0^0}{d\cos \vartheta} = C$$

$$\widetilde{Y}_0^0 = C \int \frac{d\cos \vartheta}{1 - \cos^2 \vartheta}$$

It becomes infinite at $\vartheta = 0$ and $\vartheta = \pi$.

Consequently, if there exists a regular solution for (29.4), it coincides with Y_l^0 up to a constant factor.

We define the function $P_l(u)$ as

$$P_l(u) = \frac{1}{2^l l!} \frac{d^l}{du^l} (u^2 - 1)^l \quad (29.5)$$

Let us verify whether it satisfies equation (29.4). We take the expression

$$y = (u^2 - 1)^l \quad (29.6)$$

which satisfies

$$(1 - u^2) \frac{dy}{du} + 2luy = 0 \quad (29.7)$$

Differentiating this relation $l + 1$ times with respect to u, we get

$$\left(\frac{d}{du}\right)^{l+1} (1 - u^2) \frac{dy}{du} + 2l \left(\frac{d}{du}\right)^{l+1} uy$$

$$= (1 - u^2) \frac{d^2}{du^2} \left(\frac{d}{du}\right)^l y -$$

$$-2\,(l+1)\,u\,\frac{d}{du}\left(\frac{d}{du}\right)^{l}y-l\,(l+1)\left(\frac{d}{du}\right)^{l}y$$

$$+2lu\frac{d}{du}\left(\frac{d}{du}\right)^{l}y+2l\,(l+1)\left(\frac{d}{du}\right)^{l}y$$

$$=(1-u^2)\frac{d^2}{du^2}\left(\frac{d}{du}\right)^{l}y-2u\,\frac{d}{du}\left(\frac{d}{du}\right)^{l}y$$

$$+l\,(l+1)\left(\frac{d}{du}\right)^{l}y$$

$$=\left(\frac{d}{du}\,(1-u^2)\,\frac{d}{du}+l\,(l+1)\right)\left(\frac{d}{du}\right)^{l}y=0 \qquad (29.8)$$

Thus, if we put $u=\cos\vartheta$, the expression $(d/du)^{l}\,y$ will satisfy the same equation (29.4) as $Y_l^0\,(\cos\vartheta)$. The numerical factors multiplying Y_l^0 and in the definition (29.5) have been so chosen as to obtain the exact equality $Y_l^0=P_l\,(\cos\vartheta)$ (see Exercise 1). The polynomials (29.5) are known as the *Legendre polynomials*.

To round out the picture, we present without proof the expression for spherical functions for $k\neq 0$:

$$[(-1)^{k}\,Y_l^{-k}\,(u,\;\varphi)]^{*}=Y_l^{k}\,(u,\;\varphi)$$

$$Y_l^{k}=\frac{(-1)^{k}}{2^{l}l\,!}\,\frac{(l+k)\,!}{(l-k)\,!}\,(1-u^2)^{k/2}\left(\frac{d}{du}\right)^{l+k}(u^2-1)^{l}\,e^{ik\varphi} \qquad (29.9)$$

Let us now find the normalization constant for the Legendre polynomials P_l from the condition

$$1=a_l^2\int_{-1}^{1}P_l^2\,(u)\,du$$

Substituting the expression (29.5) and integrating l times by parts, we obtain

$$1=a_l^2\,\frac{1}{(2^{l}l\,!)^2}\,(-1)^{l}\int_{-1}^{1}(u^2-1)^{l}\left(\frac{d}{du}\right)^{2l}(u^2-1)^{l}\,du$$

All the integrated terms vanish at the limits, because the order of the derivative in one of the factors in them is less than l, and after differentiation there remains a certain power of (u^2-1) not equal to zero. The $(2l)$th derivative of $(u^2-1)^{l}$ is equal to $(2l)!$. Thus we arrive at the normalization condition:

$$a_l^2\,\frac{(2l)\,!}{(2^{l}l\,!)^2}\int_{-1}^{1}(1-u^2)^{l}\,du=1$$

The integral here is called an *Euler integral of the second kind*. Substitution of $(1+u)/2=v$ reduces it to standard form:

$$\int_{-1}^{1}(1-u^2)^{l}\,du=2^{2l+1}\int_{0}^{1}v^{l}\,(1-v)^{l}\,dv$$

and is equal to

$$\frac{2^{2l+1}(l\,!)^2}{(2l+1)\,!}$$

We thus find the normalization coefficient a_l:

$$a_l = \left(\frac{2l+1}{2}\right)^{1/2} \tag{29.10}$$

For $k \neq 0$ the normalization coefficient is

$$a_{lk} = \left(\frac{2l+1}{2}\right)^{1/2} \left(\frac{(l-k)\,!}{(l+k)\,!}\right)^{1/2} \tag{29.11}$$

Radial Functions. The equations for radial functions are investigated as follows. We begin by substituting the wave function in the form

$$\psi = \chi(r)\, Y_l^k (\cos\vartheta,\ \psi) \tag{29.12}$$

into the Schrödinger equation $\hat{\mathscr{H}}\psi = E\psi$, where $\hat{\mathscr{H}}$ is given by (29.1). The operator $\hat{M}^2$ involved in $\hat{\mathscr{H}}$ in operating on Y_l^k yields a constant number $h^2 l\,(l+1)$. The partial derivative with respect to r applied to $\chi(r)$ should be replaced by the total derivative. As a result we obtain the equation for the radial function:

$$-\frac{h^2}{2m}\frac{1}{r^2}\frac{d}{dr}r^2\frac{d\chi}{dr} + \frac{h^2 l\,(l+1)}{2mr^2}\chi + U(r)\chi = E\chi \tag{29.13}$$

It is convenient to reduce this equation to one-dimensional form by means of the substitution

$$\chi = \frac{R}{r} \tag{29.14}$$

Without repeating the computations we used to transform Eq. (20.5), we write the final equation for finding the energy eigenvalues:

$$-\frac{h^2}{2m}\frac{d^2R}{dr^2} + \frac{h^2 l\,(l+1)\,R}{2mr^2} + U(r)\,R = ER \tag{29.15}$$

As long as the form of $U(r)$ has not yet been made definite, we can consider (29.15) only in two limiting cases: for very large and for very small distances from the nucleus.

The field of the atomic core is not effective at very small distances from the nucleus, and there remains only the Coulomb law $U(r) = -Ze^2/r$ (Z is the atomic number of the element). However, if r is very small, the term $[h^2 l\,(l+1)/(2mr^2)]\,R$ is, in any case, larger than the term $-(Ze^2/r)\,R$, and all the more greater than ER. For the time being we put $l \neq 0$. Hence, in direct proximity to the nucleus the wave equation is of a very simple form:

$$\frac{d^2R}{dr^2} = \frac{l\,(l+1)\,R}{r^2} \tag{29.16}$$

In this form it is solved by the substitution

$$R = r^\alpha \qquad (29.17)$$

so that

$$\alpha\,(\alpha - 1) = l\,(l + 1) \qquad (29.18)$$

This equation has two roots,

$$\alpha = l + 1 \quad \text{and} \quad \alpha = -l \qquad (29.19)$$

But from (29.14), the second root yields $\chi = r^{-l-1}$; at the point $r = 0$ such a function χ becomes infinite for all l. Therefore, we must discard the root $\alpha = -l$ and take the dependence of R on r for small l's in the form

$$R = Cr^{l+1}, \qquad \chi = Cr^l \qquad (29.20)$$

For $l = 0$ the Coulomb term must be retained in the equation. This equation is solved by substituting R in the form of a series

$$R = a_1 r + a_2 r^2 + \ldots \qquad (29.21)$$

so that the dependence of the wave function upon the radius at small distances from the nucleus has the form (29.20) for all l's.

The greater the angular momentum, the higher the order of the wave-function's zero at the coordinate origin. Only for $l = 0$ does it remain finite close to the nucleus: $\chi = R/r = a_1$. This can be understood by analogy with classical mechanics: angular momentum is the product of momentum by the "arm", i.e., by the distance from the origin; $l = 0$ corresponds to a zero "arm" and a zero angular momentum. Therefore, there is a nonzero probability of finding the electron at the origin. In the old version of quantum mechanics (due to Bohr) the electron orbit with zero angular momentum passed through the nucleus, which could not be explained. The larger angular momentum values correspond to larger "arms" and, correspondingly, in quantum mechanics, to a smaller probability of finding an electron close to the nucleus.

The behaviour of the wave function close to the origin can also be explained as follows. A repulsive centrifugal force acts on the particle; to this force there corresponds an effective potential $\hbar^2 l\,(l + 1)/(2mr^2)$ (see Sec. 5). This limits the classically possible region of motion for small r's. In quantum mechanics the particle penetrates the region where according to classical mechanics the velocity would have to be imaginary. But the wave function decreases rapidly with penetration into this region, that is, as it approaches the origin of the coordinate system. The decrease in this case is according to the power law: as $r \to 0$, the wave function decreases as r^l, that is, the damping law is the stronger the higher the barrier. At $l = 0$ there is no barrier, and nothing prevents the electron from drawing infinitely close to the nucleus.

Let us now examine the region of large values of r. For large r's, in Eq. (29.15) we can discard the terms that decrease as r increases, that is, the centrifugal and Coulomb energy (the latter is gauged to zero at infinity: $U(\infty) = 0$). Then the equation is greatly simplified:

$$\frac{d^2R}{dr^2} = -\frac{2mE}{h^2} R \tag{29.22}$$

Its general solution appears thus:

$$R = C_1 \exp\left(-\frac{(-2mE)^{1/2}}{h} r\right) + C_2 \exp\left(\frac{(-2mE)^{1/2}}{h} r\right) \tag{29.23}$$

(i) Let the energy be positive, $E > 0$. Here, R appears as follows:

$$R = C_1 \exp\left(\frac{i(2mE)^{1/2}}{h} r\right) + C_2 \exp\left(-\frac{i(2mE)^{1/2}}{h} r\right) \tag{29.24}$$

Both terms remain finite for any value of r. Therefore, the two constants C_1 and C_2 must be retained in the solution. We came across the same situation in Section 28 in considering the solution of the wave equation (28.33) for a potential well of finite depth.

Any general solution of a second-order differential equation involves two arbitrary constants. Let us suppose that the solution (29.20), which holds for small r's only, is continued into the region of large r's, where it is not of the simple form r^{l+1} but nevertheless satisfies the precise equation (29.15). A certain integral curve is thus obtained for this equation. But any integral curve can be represented by properly choosing the two constants in the general solution. As r tends to infinity this solution acquires its asymptotic form (29.24), if $E > 0$. The expression (29.24) remains finite when $r \to \infty$ for any constants C_1 and C_2. It follows that, for a positive energy, the wave equation always has a finite solution for any values of r. Therefore, the energy region $E > 0$ corresponds to a continuous spectrum, since the wave function satisfies the required conditions at zero and at infinity for any $E > 0$. In accordance with (29.24), the probability of finding an electron for $r \to \infty$ does not become zero; that is, this case corresponds to infinite motion, as in the classical problem considered in Section 5 (see also Sec. 28).

Thus, the general rule that infinite motion possesses a continuous energy spectrum has been confirmed.

(ii) Now let $E < 0$, or $E = -|E|$. Then (29.23) must be represented thus:

$$R = C_1 \exp\left(\frac{(2m|E|)^{1/2}}{h} r\right) + C_2 \exp\left(-\frac{(2m|E|)^{1/2}}{h} r\right) \tag{29.25}$$

24*

Here the first term tends to infinity with r, and we must therefore put $C_1 = 0$, so that R involves one arbitrary constant instead of two:

$$R = C_2 \exp\left(-\frac{(2m\,|\,E\,|\,)^{1/2}}{h}\, r\right) \tag{29.26}$$

If we now draw an integral curve from the coordinate origin, starting from $r = 0$, where it has the form (29.20), for large r, as a rule, it will not be reduced to the form (29.26). For all negative energy values, except certain ones, the integral curve is represented in the form (29.25) at infinity and, hence, does not satisfy the boundary condition imposed on the wave function. Only for those energy values for which the plot of integral curve is such that

$$C_1(E) = 0 \tag{29.27}$$

does the wave equation have a solution satisfying the boundary conditions. This corresponds to a discrete energy spectrum. At the same time, $R(\infty)$ becomes zero, so that the finite motion has a discrete energy spectrum, as expected.

The Coulomb Field. Let us now see how a discrete energy spectrum is found in the case of a purely Coulomb field:

$$U(r) = -\frac{Ze^2}{r} \tag{29.28}$$

This occurs in a hydrogen atom (though not in a molecule!), in singly ionized helium, doubly ionized lithium, etc.

The wave equation (29.15) is now written as

$$-\frac{h^2}{2m}\frac{d^2R}{dr^2} + \frac{h^2 l\,(l+1)}{2mr^2}\,R - \frac{Ze^2}{r}\,R = -|E|\,R \tag{29.29}$$

where we have straightaway taken the case of negative energies, which leads to a discrete spectrum.

It is convenient here to change the units of length and energy similar to the way it was done in the problem of the harmonic oscillator (Sec. 28). In place of the system where the basic units are the arbitrary quantities centimetre, gram, second we take the following units: the elementary charge e, the mass of the electron m, and the quantum of action h. From these quantities we form the unit of length:

$$\frac{h^2}{me^2} = (5.29172 \pm 0.00002) \times 10^{-9}\ \text{cm}$$

and the unit of energy:

$$\frac{me^4}{h^2} = (27.20976 \pm 0.00044)\ \text{eV}$$

(the units of length and energy are expressed in terms of the actual mass of the electron; in a hydrogen atom we must take the reduced electron-proton mass).

Hence, if we put $e = 1$, $m = 1$, $h = 1$ in Eq. (29.29), then length and energy is measured in these units. Let us denote this length ξ:

$$\xi = \frac{me^2}{h^2}\, r \tag{29.30}$$

and energy ε:

$$\varepsilon = \frac{h^2}{me^4}\, |E| \tag{29.31}$$

so that, of the constants, the wave equation will involve only the atomic number Z:

$$-\frac{d^2R}{d\xi^2} + \frac{l(l+1)}{\xi^2}\, R - \frac{2Z}{\xi}\, R = -2\varepsilon R \tag{29.32}$$

We look for the solution of this equation in the form of a series expansion. We shall proceed here from the solutions obtained for large and small values of ξ (or r).

In accordance with Eqs. (29.20) and (29.23), we write R in the following form:

$$R = \xi^{(l+1)} e^{-\xi(2\varepsilon)^{1/2}} (\chi_0 + \chi_1 \xi + \chi_2 \xi^2 + \cdots)$$
$$= \xi^{(l+1)} e^{-\xi(2\varepsilon)^{1/2}} \sum_{n=0}^{\infty} \chi_n \xi^n = e^{-\xi(2\varepsilon)^{1/2}} \sum_{n=0}^{\infty} \chi_n \xi^{n+l+1} \tag{29.33}$$

The first factor determines the form of R as $\xi \to 0$, the second factor should basically correspond to the form of R for large ξ's, and the series interpolates, as it were, between the limiting values. Differentiating (29.33) twice, we obtain

$$\frac{d^2R}{d\xi^2} = 2\varepsilon e^{-\xi(2\varepsilon)^{1/2}} \sum_{n=0}^{\infty} \chi_n \xi^{n+l+1}$$

$$- 2(2\varepsilon)^{1/2} e^{-\xi(2\varepsilon)^{1/2}} \sum_{n=0}^{\infty} (n+l+1)\chi_n \xi^{n+l}$$

$$+ e^{-\xi(2\varepsilon)^{1/2}} \sum_{n=0}^{\infty} (n+l+1)(n+l)\chi_n \xi^{n+l-1}$$

The first term on the right is simply $-2\varepsilon R$. Hence, it cancels with the same term in (29.32) on the right. We group the remaining terms so that in one of them the degree of ξ is everywhere less by unity than in (29.33) and, in the other, less by two units. In addition we eliminate the exponential factor. We shall now have an equality between two such series:

$$\sum_{n=0}^{\infty} [l(l+1) - (n+l+1)(n+l)] \chi_n \xi^{n+l-1}$$

$$= \sum_{n=0}^{\infty} [2Z - 2(2\varepsilon)^{1/2}(n+l+1)] \chi_n \xi^{n+l}$$

Such an equality is possible only when the coefficients of the same powers of ξ coincide. In the left-hand side the power ξ^{n+l} will have a coefficient involving χ_{n+1}. Hence

$$\chi_{n+1} = \chi_n \frac{2\left[Z - (n+l+1)(2\varepsilon)^{1/2}\right]}{l(l+1) - (n+l+1)(n+l+2)} \tag{29.34}$$

From the relationship (29.34), all the χ_n's are determined consecutively. We must neglect the constant numbers Z, l and $l(l+1)$ in Eq. (29.34) in comparison with n when n's are large; there then remains the limit

$$\chi_{n+1} = \chi_n \frac{2(2\varepsilon)^{1/2}}{n} \tag{29.35}$$

We met with a similar limiting expression in the problem of the linear harmonic oscillator (Sec. 28). In the case of large ξ's it reduces the whole series to an exponential form:

$$\sum_n \chi_n \xi^n \approx e^{2\xi(2\varepsilon)^{1/2}} \tag{29.36}$$

But such a series cannot give a correct expression for $R(\xi)$, because, if we substitute (29.36) into (29.33), we obtain $\psi(\infty) = \infty$ despite the boundary condition. However, if all the coefficients become zero from a certain χ_{n+1} onwards, the series (29.33) becomes a polynomial. Then, being multiplied by $e^{-\xi(2\varepsilon)^{1/2}}$, it gives $\psi(\infty) = 0$, as required. It can be seen from (29.34) that χ_{n+1} vanishes if

$$Z - (n+l+1)(2\varepsilon)^{1/2} = 0 \tag{29.37}$$

or

$$\varepsilon = \frac{Z^2}{2(n+l+1)^2} \tag{29.38}$$

Finally, going over to conventional units and taking into account the sign of the energy, we obtain the required spectrum:

$$E = -\frac{Z^2 m e^4}{2h^2(n+l+1)^2} \tag{29.39}$$

Quantum Numbers. The number n is the power of the polynomial $\sum \chi_n \xi^l$ involved in the wave function expression. A more detailed analysis shows that the polynomial has exactly n real roots. Since $\chi_0 \neq 0$ and $\chi_n \neq 0$, none of the roots are equal to either zero or infinity. Therefore, if we examine the dependence of the wave function on the radius, it has n zeros, or "nodes", not counting the zero at $\xi = \infty$, associated with the finiteness of the motion, and at $\xi = 0$, which occurs in all functions with $l \neq 0$. The term "node" is used instead of "zero" by analogy with the nodes of a vibrating string fixed at both ends (the latter problem is completely analogous to the motion of a particle in an infinitely deep potential well).

Obviously, the wave function has zeros not only in the case of a hydrogen atom. Therefore, the number n_r, or the number of nodes of the function at a finite distance from the origin of the coordinate system, can characterize the state of an electron in any atom, insofar as, from the qualitative aspect, it is legitimate to describe the action of all other electrons on an individual electron with the help of the effective potential $U(r)$. For the same reason, it is possible to describe the state of a separate electron with the help of the number l, which is used to express the square of the angular momentum. There is one more number that is expressed in terms of n_r and l; it is denoted n and related to n_r and l as follows:

$$n \equiv n_r + l + 1 \qquad (29.40)$$

These quantities can also be used for complex, many-electron atoms, even though the energy of the electrons in them cannot be expressed by the simple formula (29.39). The numbers n, n_r, and l are convenient for classifying electron states.

The number l is called the *orbital quantum number* of an electron. In spectroscopy the following system of notation is accepted: an electron state with $l = 0$ is called the s state, with p, d, and f states corresponding to $l = 1$, 2, and 3. Higher values of l do not occur in unexcited atoms. The total angular momentum of an atom as a whole is found by adding the angular momenta of individual electrons (see following section).

As we know, the projection of the angular momentum of a separate atom on some axis is equal to hk, where $-l \leqslant k \leqslant l$. The integer k is called the *magnetic quantum number*, since as a rule reference is to the axis along which an external magnetic field is directed.

Next, n_r is the number of zeros (nodes) in the radial wave function, and it is called the *radial quantum number*.

Finally, the sum (29.40), n, is called the *principal quantum number* of an electron in an atom.

From (29.39) and (29.40), the energy of an electron in a hydrogen atom expressed in terms of the principal quantum number is

$$E_n = -\frac{me^4}{2h^2n^2} = -\frac{13.5}{n^2} \text{ eV} \qquad (29.41)$$

If an external source imparts this energy to the electron in a hydrogen atom, the electron may be ejected from the atom. For this reason the corresponding energy is known as the *binding energy*. A formula analogous to (29.41) is obtained also for the positive helium ion. Apart from the $Z^2 = 4$-fold difference, there is a more subtle distinction stemming from a slight difference between the reduced mass of the helium atom and the reduced mass of the hydrogen atom due to the differences between masses made up of nuclei

(we recall that the reduced mass of an atom with one electron closely approximates the mass of the electron).

The state with $n = 1$ is the ground state. The atom cannot emit light in this state, because there is simply no lower state into which it can pass.

We pointed out at the beginning of this section that the energy of a particle moving in a central field can depend only on the square of its angular momentum, but not on the latter's projection. In other words, wave functions corresponding to identical l's and, as we now know, identical n_r's but different k's correspond to the same electron energy. This is what is called degeneracy.

We find a somewhat unexpected result for the Coulomb field: the energy depends not on the two quantum numbers n_r and l but only on their sum $n_r + l$. For different n_r and l, but such that their sum is the same, the energy eigenvalues are also the same. Consequently a further degeneracy of energy states occurs, since to different l's correspond quite different spherical functions Y_l^k.

Note should be made of a substantial difference between degeneracy with respect to k and with respect to l. The first occurs for any analytical form of the potential $U(r)$ and is connected only with the symmetry of the force field acting on the particle. In the case of $U = U(r)$, it is symmetry of all rotations around the origin of the coordinate system. This type of degeneracy is called *necessary* and is due to the symmetry specified in the conditions of the problem on the determination of energy eigenvalues.

Degeneracy in a Coulomb field, however, is due wholly to the specific form of the dependence $U = -Ze^2/r$. If, for example, the potential energy curve is of exponential form, $U = -a/r^b$, degeneracy with respect to l occurs only when the exponent in the force law is $b = 1$. This type of degeneracy is called *accidental*: it occurs at some select value of the parameter involved in the Hamiltonian.

There is a deep connection between the accidental degeneracy of states in a Coulomb field in classical and quantum mechanics. In classical mechanics (Sec. 10) it was shown that in Keplerian motion energy depends only upon the sum of the adiabatic invariants J_r and J_φ, but not on each of them separately. Subsequently (Sec. 31) it will be shown that there is a very simple relationship between adiabatic invariants and quantum numbers. Therefore, Eq. (29.39) is a direct quantum analog of the classical formula expressing the dependence of energy on the adiabatic invariants (see Exercise 3, Section 10).

In turn, the classical form of the dependence of energy upon the sum of the adiabatic invariants J_r and J_φ determines the closed nature of the path of finite motion in the Kepler problem (an ellipse). In the relativistic Kepler problem (Exercise 9, Section 14) the

path is not closed, and the dependence of the energy upon the adiabatic invariants is more complex. Correspondingly, in quantum theory, too, there is no accidental degeneracy with respect to l.

Parity of State. The state of an electron in an atom is characterized by one more property which, unlike energy and angular momentum, has no classical analogue. This property relates directly to the wave function itself and is associated with its behaviour under changes of the signs of all three coordinates (cf. Sec. 15).

Consider the wave function of an electron. The form of the wave equation (29.15) does not change if we substitute

$$x = -x', \qquad y = -y', \qquad z = -z' \qquad (29.42)$$

As we know from Section 15, this transformation is called *inversion*, and it transforms a right-handed coordinate system into a left-handed one, and vice versa. No spatial rotation of the axes can make the two systems coincide.

The wave equation (29.15) is linear. Therefore, if it has not changed its form, its solution, determined by the boundary conditions up to a constant factor, can acquire only a certain additional factor:

$$\psi (x, y, z) = C\psi (x', y', z') = C\psi (-x, -y, -z) \quad (29.43a)$$

But, in principle, the primed left-handed system in no way differs from the unprimed right-handed system. Hence, the reverse transformation must involve the same transformation factor C:

$$\psi (x', y', z') = \psi (-x, -y, -z) = C\psi (x, y, z) \qquad (29.43b)$$

Substituting this into (29.43a), we obtain

$$\psi (x, y, z) = C^2\psi (x, y, z)$$

whence

$$C^2 = 1, \qquad C = \pm 1 \qquad (29.44)$$

The function is said to be *even* for $C = 1$ and *odd* for $C = -1$. The wave functions of a linear harmonic oscillator possessed a similar property: the energy operator was also even, $\hat{\mathcal{H}}(x) = \hat{\mathcal{H}}(-x)$, while the parity of the wave functions alternated, depending on the eigenvalue number n, that is, they were either even or odd.

It is not hard to establish the quantum number that determines the parity of an electron's wave function in a central field. From the spherical function expression (24.39), it is obtained by differentiating the function r^{-1} with respect to the coordinates l times. Consequently, the number l fully defines the parity of the wave function of a particle in its coordinate dependence.

If a many-electron atom is considered, the parity of its total wave function is equal to the parity of the number $\sum l_i$, where l_i are the orbital quantum numbers of the individual electrons.

The parity of a wave function can be represented in operator form by introducing an inversion operator $\hat{G}$ such that

$$\hat{G}\psi(x, y, z) = \psi(-x, -y, -z) \tag{29.45}$$

Since the Hamiltonian for an atom is an even function of the coordinates, we can write

$$\hat{G}\hat{\mathscr{H}} = \hat{\mathscr{H}} \tag{29.46}$$

From this it follows that the parity operator commutes with the Hamiltonian:

$$\hat{G}\hat{\mathscr{H}}\psi = \hat{\mathscr{H}}\hat{G}\psi \tag{29.47}$$

The eigenvalues of the operator $\hat{G}$ are, in accordance with (29.44), the numbers $C = \pm 1$, since

$$\hat{G}\psi = \psi(-x, -y, -z) = C\psi \tag{29.48}$$

According to (29.47) and (27.8), these numbers exist simultaneously with the energy eigenvalue.

Speaking of a separate electron, giving its parity of state provides no new information, since parity is determined by the orbital quantum number l. In a many-electron atom the state is given not only by the quantum numbers of individual electrons, but also by how the angular momenta of individual electrons add up into the resultant angular momentum of the atom as a whole.

Addition of Angular Momenta. To begin with, we shall consider the rule for the addition of the angular momenta of two electrons in an atom. The angular momentum of each electron is not, strictly speaking, an integral of the motion, since the electron moves in the field not only of the nucleus but of the other electron as well. Such a field does not possess symmetry with respect to rotations about the nucleus, therefore only the total angular momentum of both electrons can be conserved, but not the angular momentum of each one separately.

Nevertheless, from the qualitative aspect, the electrons' orbital quantum numbers l_1 and l_2 continue to provide a correct description of their respective states. They help to define the exact quantum number of the total angular momentum.

We shall reason as though l_1 and l_2 were also exact quantum numbers. But having determined the total angular momentum with their help, we must bear in mind that only it is an exact quantum integral of motion of the system.

To make the reasoning specific, let $l_1 > l_2$. We project the smaller angular momentum on the direction of the greater. Since the projection of l_2 varies from l_2 to $-l_2$, we find that in sum with the greater angular momentum l_1 we obtain the following possible maximum projections of the total angular momentum (in units of h):

$$L = l_1 + l_2, \quad l_1 + l_2 - 1, \quad l_1 + l_2 - 2, \ldots, l_1 - l_2$$

$$(29.49)$$

Each of these values defines the square of the resultant angular momentum according to the formula

$$M_{total}^2 = h^2 L (L + 1) \tag{29.50}$$

It is not difficult to derive this formula by the method of mean values in the same way as we developed the formula for the angular momentum square of a separate electron (Exercise 1, Section 25).

Thus, the angular momentum of two electrons varies within the limits from $L = l_1 + l_2$ to $L = |l_1 - l_2|$. The projection of the resultant angular momentum L may vary from $-L$ to L. The rule set forth here agrees with the fact that the magnitude of the sum of two vectors lies between the sum and the difference of their absolute values.

By analogy with the states of individual electrons labelled s, p, d, f for $l = 0, 1, 2, 3$, the states of an atom are denoted by capital letters S, P, D, F, ... for $L = 0, 1, 2, 3, \ldots$. To larger L's correspond the letters following F in alphabetical order.

The generalization of the rule presented above for the case of three or more electrons is self-evident: first any two angular momenta are added according to the rule (29.49), then a third is added according to the same rule, etc.

The Simultaneous Operation of the Angular Momentum and Parity Conservation Laws. We have thus established that a system of electrons in the central field of a nucleus is subject to two conservation laws: conservation of total angular momentum and of total parity. Unlike the case of one electron in a system of electrons, these two laws by no means reduce to the same thing. The total parity is found by arithmetic addition of the individual numbers l, while the total angular momentum is determined by geometrical (vector) composition. Therefore, to describe the state of a system of electrons we must define the angular momentum and parity corresponding to that state.

Let us now examine the restrictions that can be imposed on the possible transitions between different atomic states by these two conservation laws operating together. As an example we take an excited many-electron atom with total angular momentum $L = 0$,

that is, in the S state. The atom has s-electrons and, we assume, an odd number of p-electrons. Consequently, the atom is in odd state (its parity is determined by the addition of an odd number of units[6]). Furthermore, let the total excitation energy of the atom be sufficient to eject one p-electron, but such that as a result of the rearrangement of the electron cloud the atom or ion remains again in the S state, with $L = 0$. Since the angular momenta are added vectorially, such a state may occur for either an even or odd number of p-electrons.

The condition that the total angular momentum must be conserved requires that the electron emitted from the atom have $l = 0$, since the total angular momentum of the initial system was zero, and after the emission of the electron there remains a system whose angular momentum is by definition zero. In other words, the assumption has been made that this is the only state in which the electron possesses sufficient energy for emission. Hence, the angular momentum conservation law requires that the angular momentum of the emitted electron also be zero.

These considerations show that, given the appropriate initial assumptions, the laws of conservation of energy and angular momentum can be satisfied. Let us see if the parity conservation law also holds in these circumstances. The remaining system now has two p-electrons with $l = 1$, so that its state is even. The emitted electron has, in accordance with the angular momentum conservation law, $l = 0$, that is, its state is also even. Consequently, the ultimate state of the system must also be even, whereas its initial state was assumed to be odd. It follows that, by the parity conservation law, the transition considered is impossible. The energy and angular momentum conservation laws, which have classical analogues, do not preclude such a transition, which is impossible according to the quantum law of conservation of parity, for which there is no classical analogue. We have examined a typical case of a transition "forbidden" for considerations of parity (from $L = 0$ to $L = 0$ with an assumed change in parity).

We repeat that the law of conservation of parity is independent of the law of conservation of angular momentum, because parity is determined by a different law of addition than angular momentum.

In quantum mechanics, the angular momentum conservation law must always be applied together with the parity conservation law. Common to both these laws is that they derive from the invariance of the equations with respect to the spatial orientation of the coordinate axes. But the orientation of the axes can be changed not only

---

[6] The vector addition of three unit angular momenta may yield zero in the following way. Two vectorially added momenta may yield an angular momentum equal to unity, since their resultant momentum varies between 0 and 2. The resultant unity may yield zero when added to $l = 1$ for the third electron.

by rotations: an additional transformation is inversion, which is not reducible to any rotation. This is what the parity conservation law provides for in addition to the angular momentum conservation law.

In this form the parity conservation law is unconditionally applicable to systems in which electromagnetic interactions occur. This is an experimental fact on the basis of which the equations of electrodynamics are made invariant with respect to inversions of the coordinate system (Sec. 15).

The much weaker interactions occurring in some elementary particle transformations (for example, β-decay) do not satisfy the parity conservation law.

Hydrogenlike Atoms. At the beginning of this section it was pointed out that alkali-metal atoms somewhat resemble the hydrogen atom. The outer electron in these atoms is relatively weakly bound to the atomic core, which consists of the nucleus and all the remaining electrons. The wave functions for electrons of the atomic core differ from zero at smaller distances from the nucleus than the wave function for the outer electron, so that the core, as it were, screens the nuclear charge. The field in which the outer electron moves is approximately Coulomb, provided only that it is not situated in the region of the core. It is for this reason that the spectra of alkali-metal atoms resemble the hydrogen-atom spectrum. The energy levels of these atoms, which are due to excitation of the outer electron, are given by the equation

$$E_{nl} = -\frac{me^4}{2h^2}\frac{1}{[n+\Delta(l)]^2} \tag{29.51}$$

where the correction $\Delta(l)$ depends upon the orbital quantum number. It accounts for the deviation of the field from a purely Coulomb one at small distances from the nucleus. The greater l is, the farther from the nucleus is the outer electron (according to (29.20)) and the smaller is $\Delta(l)$.

Thus, the energy levels of alkali metals, like the energy levels of all atoms except hydrogen, depend upon n and l.

EXERCISES

1. Show that in the definition (29.3) the spherical function coincides fully with the Legendre polynomial (29.5) of $u = \cos\vartheta$.

Solution. Since it has already been shown that the functions (29.3) and (29.5) are proportional, it is sufficient to prove their identity for some value of $u = \cos\vartheta$. We shall show this for $\cos\vartheta = u = 1$. In the definition (29.5), at $u = 1$ only the term that is obtained from l-fold differentiation

of the binomial $(u^2 — 1)^l$ is not zero. Each differentiation yields a factor equal to the power of the binomial, and a 2 from the differentiation of the square. These factors cancel out with the denominator $2^l l!$, so that $P_l(1) = 1$. Now let us prove that $Y_l^0(1)$ is also unity.

In the definition (29.3) we shift the origin of the coordinate system by the quantity ζ in the direction of the z axis so as to make r^{-1} under the derivative sign equal to $[x^2 + y^2 + (z — \zeta)^2]^{-1/2}$. For small ζ's we may leave r under the derivative sign equal to $[x^2 + y^2 + z^2]^{1/2}$.

We divide both sides of the equation by r^l, multiply by ζ^l, put $\zeta = 0$ in the expression for the derivative, and sum over l from 0 to ∞. Then in the right-hand side we have a Taylor series of the quantity $[x^2 + y^2 + (z — \zeta)^2]^{-1/2}$.

Hence, $Y_l^0(\cos \vartheta)$ is the expansion coefficient of $(\zeta/r)^l$ in the Taylor series of $[x^2 + y^2 + (z — \zeta)^2]^{-1/2}$:

$$r\,[x^2+y^2+(z—\zeta)^2]^{-1/2} = \left[1 - 2\frac{\zeta}{r}\cos\vartheta + \frac{\zeta^2}{r^2}\right]^{-1/2}$$

$$= \sum_{l=0}^{\infty} \left(\frac{\zeta}{r}\right)^l Y_l^0(\cos\vartheta)$$

But for $\cos\vartheta = 1$ we have simply

$$\left[1 - 2\frac{\zeta}{r} + \frac{\zeta^2}{r^2}\right]^{-1/2} = \left(1 - \frac{\zeta}{r}\right)^{-1} = \sum_{l=0}^{\infty}\left(\frac{\zeta}{r}\right)^l$$

Hence, the required coefficient is equal to unity, as was asserted.

2. Prove that three successive Legendre polynomials are related by the formula

$$(2l+1)\,uP_l(u) = (l+1)\,P_{l+1}(u) + lP_{l-1}(u)$$

Solution. Using the result of the preceding exercise, we write

$$(1 - 2\rho u + \rho^2)^{-1/2} = \sum_{l=0}^{\infty}\rho^l P_l(u)$$

Differentiating this equation with respect to ρ, multiplying by 2ρ, and adding the initial expression, we obtain

$$\sum_l (2l+1)\,\rho^l P_l(u) = (1 - \rho^2)(1 - 2\rho u + \rho^2)^{-3/2}$$

Substituting $l + 1$ for the summation index l in the derivative with respect to ρ, we obtain

$$\sum_l (l+1)\,\rho^l P_{l+1}(u) = (u - \rho)(1 - 2\rho u + \rho^2)^{-3/2}$$

Now we perform the following operations: multiply the derivative with respect to ρ by ρ^2, add the initial expression multiplied by ρ, and

replace the summation index l by $l-1$. We then obtain

$$\sum_{l=0}^{\infty} l\rho^l P_{l-1}(u) = (\rho - \rho^2 u)(1 - 2\rho u + \rho^2)^{-3/2}$$

We see that the sum of the two latter equations coincides with the first, multiplied by u. Comparing the factors of the same powers of ρ, we arrive at the required equation.

It is apparent that this equation could not involve a polynomial of order higher than $l + 1$, because the power of P_l is equal to l. Furthermore, since multiplication by u changes the parity of P_l, the equation holds only for polynomials of different parity on the left and on the right.

3. Develop and normalize the wave functions in a hydrogen atom with $l = 0$, 1, 2 and $n = 1$, 2, 3. Take advantage of the fact that

$$\int_0^{\infty} e^{-x} x^n \, dx = n!$$

30

ELECTRON SPIN

From Eq. (29.41) the ground state of a hydrogen atom has the principal quantum number, n, equal to unity. For $n = 1$ the orbital quantum number, l, and the radial quantum number, n_r, must be equal to zero, since $n = n_r + l + 1$, and n_r and l can in no way be less than zero. The ground state of a hydrogen atom is thus the s state. The orbital motion of an s-electron does not produce a magnetic moment because the magnetic moment is proportional to the angular momentum. Yet, if the Stern-Gerlach experiment is performed for atomic hydrogen, the atomic beam will split, but only into two parts. However, when $l = 0$, as we have already said, there should be no splitting due to orbital angular momentum, while for $l = 1$, the beam should split into 3 beams corresponding to the number of projections of the angular momentum $k = -1$, 0, 1.

The same results if, instead of hydrogen, we take an alkali metal. The electron cloud of any alkali metal consists of an atomic core in the S state, that is, one lacking orbital angular momentum, and one electron in the s state. In this sense, alkali-metal atoms resemble the hydrogen atom.

For this reason, the state of the atom is not described by the three quantum numbers n, l, and k. One more quantum number must be stated, with respect to which splitting of the beam occurs.

Intrinsic Angular Momentum, or Spin, of an Electron. Obviously, the additional quantum number must be associated in some way with the angular momentum of the electron, since it leads to a splitting of the beam in a magnetic field. Splitting into two beams can be accounted for only by an angular momentum whose greatest projection is equal to $h/2$. Then it has only two possible projections, $h/2$ and $-h/2$.

The Stern-Gerlach experiment was given only as an example. In fact, not only this experiment, but the whole enormous aggregate of knowledge about the atom indicates that the electron possesses an angular momentum $h/2$ that is not related to its spatial, or as it is conventionally called, orbital, motion. This angular momentum is termed the *spin*. It can be said that in the classical analogy an electron is like a planet, which has an angular momentum due not only to its revolution about the sun, but also to rotation on its own axis.

The analogy with a planet is not far-reaching since the angular momentum of a rotating rigid body can be made equal to any value, while the spin of an electron always has projections $\pm h/2$ and no others. Therefore, spin is a purely quantum property of the electron; in the limiting transition to classical mechanics it becomes zero. We must not take the word "spin" too literally, for the electron actually does not resemble a rigid body like a top or a spindle. The analogy between an electron and a top consists only in that their motion is not described solely by the spatial location of one point, and they possess an internal rotational degree of freedom.

There is, rather, a certain analogy between the electron and the quantum of light: as was shown in Section 19, an electromagnetic wave possesses an internal polarization degree of freedom. Two waves of identical phase may possess different polarization. If we liken the coordinate-dependent phase with the spatial argument of the wave function of the electron, the polarization degree of freedom of the wave can be likened to the spin degree of freedom of the quantum. But the two are far from identical: wave polarization is a classical concept, while spin is a quantum concept.

The General Definition of Angular Momentum. Since spin is not associated with the spatial motion of the electron, definition of the angular momentum due to it according to the operator formula $\hat{\mathbf{M}} = \hat{\mathbf{r}} \times \hat{\mathbf{p}}$ is, evidently, unacceptable. A more general definition is required, which would be valid for all cases, especially, of course, one that would describe the property of angular momentum as an integral of the motion.

In classical mechanics, the angular momentum of a mechanical system is defined (for one projection) as

$$M_z = \frac{\delta S}{\delta \varphi} \tag{30.1}$$

Here, S denotes the action of the system, and $\delta \varphi$ is an infinitesimal angle of rotation about the z axis. The rotation $\delta \varphi$ stresses the fact that it possesses a somewhat different meaning than in the previous definition of angular momentum (see Secs. 5 and 24): φ denotes not the azimuth angle of a separate mass point, but the angle of rotation of a rectangular coordinate system. In the present case the rotation is infinitesimal.

If we take a system of mass points, it is obvious that a rotation of the coordinates through $\delta \varphi$ is equivalent to a displacement of each point by $\delta \varphi$ along the azimuth, so that in this case definition (30.1) yields the projection of the total angular momentum as an additive integral of the motion. The same is true of the angular momentum of a rotating solid body, all points of which rotate through the same angle.

Angular momentum as an additive integral of motion of a closed system exists only because there exists the symmetry of relative rotations, which we called the isotropy of space in Section 2. The general definition of angular momentum in quantum mechanics can also be linked with such symmetry, irrespective of whether the angular momentum is due to the spatial displacement of a mass point or to some internal degree of freedom that cannot be described with the help of conventional coordinates. The correctness of such a generalization is verified, as always, by experiment.

If the angular momentum is due solely to spatial motion, it is more convenient to proceed from the classical equivalence between wave function and action:

$$\psi = e^{iS/h}$$

(see (23.5)). We find the change in ψ in an infinitesimal rotation of the coordinate system:

$$\delta \psi = \delta e^{iS/h} = e^{iS/h} \left(\frac{i}{h} \right) \delta S = \psi \, \frac{i}{h} \, M_z \delta \varphi \tag{30.2}$$

where we made use of Eq. (30.1).

The generalization consists in replacing M_z by a certain angular momentum J_z of arbitrary origin. Then (30.2) must be rewritten in the form

$$\frac{h}{i} \frac{\delta \psi}{\delta \varphi} = J_z \psi \tag{30.3}$$

where ψ is the wave function, which is no longer associated with the action since it may involve other variables besides spatial ones.

We shall treat Eq. (30.3) as the quantum definition of the operator $\hat{J}$, in other words, we pass from the magnitude of the angular momentum to the operator of the angular momentum. Let us show that in this choice of the angular momentum operator its projections satisfy the conventional commutation rules (24.24), which were developed for the projections of the angular momentum of spatial motion of a particle. They are termed the orbital momentum projections (the term "orbital" goes back to Bohr's theory, which assumed the existence of orbits).

Every operator is defined with respect to some entity on which it operates: in geometrical space, for example, this may be a vector, in Hilbert space a state vector, that is, a wave function. It is convenient to begin by obtaining the commutation relation for the operators $\hat{J}_x$, $\hat{J}_y$, $\hat{J}_z$ in their operation on the radius-vector components x, y, z.

If we introduce three unit vectors $\mathbf{n}^{(1)}$, $\mathbf{n}^{(2)}$, and $\mathbf{n}^{(3)}$, then, as we know from Sections 8 and 9, the change in the radius vector in a rotation of the coordinate system is

$$\delta_i \mathbf{r} = \mathbf{n}^{(i)} \delta\varphi_i \times \mathbf{r} \tag{30.4}$$

where $\delta\varphi_i$ is an infinitesimal rotation around the ith axis.

We rotate the coordinate system once more about the kth axis $(k \neq i)$; then the change of $\delta_k \mathbf{r}$ in the new rotation is

$$\delta_k (\delta_i \mathbf{r}) = \mathbf{n}^{(k)} \delta\varphi_k \times \delta_i \mathbf{r} = [\mathbf{n}^{(k)} \times (\mathbf{n}^{(i)} \times \mathbf{r})] \, \delta\varphi_i \, \delta\varphi_k \tag{30.5a}$$

Now let us perform the same operations in reverse order:

$$\delta_i (\delta_k \mathbf{r}) = \mathbf{n}^{(i)} \delta\varphi_i \times \delta_k \mathbf{r} = [\mathbf{n}^{(i)} \times (\mathbf{n}^{(k)} \times \mathbf{r})] \, \delta\varphi_i \, \delta\varphi_k \tag{30.5b}$$

and find the difference between both changes in the radius vector:

$$\delta_i (\delta_k \mathbf{r}) - \delta_k (\delta_i \mathbf{r}) = [\mathbf{n}^{(i)} \times (\mathbf{n}^{(k)} \times \mathbf{r}) - \mathbf{n}^{(k)} \times (\mathbf{n}^{(i)} \times \mathbf{r})] \, \delta\varphi_i \, \delta\varphi_k$$
$$= [(\mathbf{n}^{(i)} \cdot \mathbf{n}^{(k)}) \, (\mathbf{r} - \mathbf{n}^{(k)} \, (\mathbf{n}^{(i)} \cdot \mathbf{r})$$
$$- (\mathbf{n}^{(i)} \cdot \mathbf{n}^{(k)}) \, (\mathbf{r} + \mathbf{n}^{(i)} \, (\mathbf{n}^{(k)} \cdot \mathbf{r})] \, \delta\varphi_i \, \delta\varphi_k$$
$$= -[(\mathbf{n}^{(i)} \times \mathbf{n}^{(k)}) \times \mathbf{r}] \, \delta\varphi_i \, \delta\varphi_k \tag{30.6}$$

But if the numbers of the unit vectors follow in cyclic order, the vector product $\mathbf{n}^{(i)} \times \mathbf{n}^{(k)} = \mathbf{n}^{(l)}$, where $i \neq k \neq l$. Consequently

$$\delta_i (\delta_k \mathbf{r}) - \delta_k (\delta_i \mathbf{r}) = -(\mathbf{n}^{(l)} \times \mathbf{r}) \, \delta\varphi_i \, \delta\varphi_k \tag{30.7}$$

In other words, rotations about different Cartesian axes do not commute. But it can be readily observed that the commutation relations for rotations about different axes are the same as the commutation relations for the respective components of angular momentum. Indeed, if we compare the rotation through angle $\delta\varphi_k$ about the kth axis with the operator $(i/h) \, \hat{J}_k \delta\varphi_k$ in the preceding equation, we

have, after cancelling out $\delta\varphi_i\delta\varphi_k$ (we assumed that $\delta\varphi_l = \delta\varphi_i\delta\varphi_k$)

$$\hat{J}_i\hat{J}_k - \hat{J}_k\hat{J}_i = ih\hat{J}_l \tag{30.8}$$

where $i \neq k \neq l$.

We have thus obtained a commutation relation for the operators of the projections of angular momentum of arbitrary origin, irrespective of the degrees of freedom to which they refer. We should assume that the commutation relations (30.8) are also valid for the operation of the operators $\hat{J}_x$, $\hat{J}_y$, and $\hat{J}_z$ in a Hilbert space of state vectors. A similar assumption was in effect made with regard to the operators of the orbital angular momentum projections, and to all operators in general.

Only on the assumption that the operators of the angular-momentum projections always satisfy the commutation relations (30.8) can angular momentum be an additive integral of the motion of a closed system. If the orbital angular-momentum projections satisfied one set of commutation relations, while the spin angular momentum projections satisfied another, then the total angular momentum would be subject to some quite special commutation relations. In particular, it would commute with the Hamiltonian differently than with the orbital angular momentum, which would violate its conservation conditions. But since experience shows that the total angular momentum of a closed system is conserved for the same conditions as those for which the orbital angular momentum alone is conserved in a system not possessing spin, it is apparent that (30.8) are the only possible commutation relations between the angular momentum components.

Note also that in a system possessing spin the strict conservation law holds only for the total angular momentum **J**, which is compounded of the orbital angular momentum and spin according to the rules of vector addition (Sec. 29). The individual angular momentum components of motion can be regarded as integrals of the motion only approximately, just as the orbital quantum numbers of separate electrons were formally treated as integrals in finding the total orbital angular momentum of the atom.

The Eigenvalues of the Angular Momentum Square and Projection. It follows from the relations (30.8), as earlier from (24.24), that angular momentum commutes with any of its projections. We shall now proceed only from the relations (30.8) to find the eigenvalues of the square of the angular momentum and one of its projections.

Take some state of a system characterized, for the time being, by an unknown eigenvalue of the square of the angular momentum, J^2, and its projection. Although we have not yet determined these numbers, we can legitimately consider the operators corresponding

to them in the state being examined to be diagonal:

$$\hat{J}^2 = J^2 \delta_{J^2 J^{2\prime}} \delta_{J_z J_z'} \tag{30.9}$$

$$\hat{J}_z = J_z \delta_{J^2 J^{2\prime}} \delta_{J_z J_z'} \tag{30.10}$$

We start with the operator expression

$$\hat{J}^2 = \hat{J}_x^2 + \hat{J}_y^2 + \hat{J}_z^2$$

from which we develop the relation between the mean values:

$$\langle J^2 \rangle = \langle J_x^2 \rangle + \langle J_y^2 \rangle + \langle J_z^2 \rangle \tag{30.11}$$

But in the eigenstates of operators the means of the corresponding quantities are equal to the eigenvalues in those states, so that

$$J^2 = J_z^2 + \langle J_x^2 \rangle + \langle J_y^2 \rangle \tag{30.12}$$

It is apparent that the mean values of the squares, $\langle J_x^2 \rangle$ and $\langle J_y^2 \rangle$, cannot be less than zero, and therefore

$$J_z^2 < J^2 \tag{30.13}$$

or

$$-\sqrt{J^2} < J_z < \sqrt{J^2} \tag{30.14}$$

We shall measure the angular momentum in natural units of h, returning when necessary to conventional units by multiplying by h. Let, then, the maximum absolute value of the angular momentum projection be J. Instead of (30.14) we write, using now the sign $\leqslant$ instead of $<$:

$$-J \leqslant J_z \leqslant J \tag{30.15}$$

We saw from the example of orbital angular momentum that $\sqrt{J^2} > J_z$. From (30.15) it follows that the matrix elements of any quantity depending upon J_z must, in the state with given J^2, vanish when even one of the indices J_z, J_z' becomes greater than J. We apply this to the matrix elements of the operators $\hat{J}_x \pm i\hat{J}_y$.

We take two commutation relations of form (30.8):

$$\hat{J}_x \hat{J}_z - \hat{J}_z \hat{J}_x = -i\hat{J}_y$$

$$\hat{J}_y \hat{J}_z - \hat{J}_z \hat{J}_y = i\hat{J}_x$$

Multiply the second by $\pm i$ and add to the first to get

$$(\hat{J}_x \pm i\hat{J}_y)\hat{J}_z - \hat{J}_z(\hat{J}_x \pm i\hat{J}_y) = \mp(\hat{J}_x \pm i\hat{J}_y) \tag{30.16}$$

Since all three operators $\hat{J}_x$, $\hat{J}_y$, and $\hat{J}_z$ commute with J^2, the respective matrices are diagonal in J^2. They have different indices

only with respect to J_z. Forming matrix elements of both sides of Eq. (30.16), we obtain

$$\sum_{J_z''=-J}^{J} [(J_x \pm iJ_y)_{J_z J_z''}(J_z)_{J_z'' J_z'} - (J_z)_{J_z J_z''}(J_x \pm iJ_y)_{J_z'' J_z'}]$$

$$= \mp (J_x \pm iJ_y)_{J_z J_z'} \quad (30.17)$$

We now take advantage of the fact that the matrix $(J_z)_{J_z J_z''}$ is itself diagonal in the representation in which J_z is the independent variable. This is expressed by Eq. (30.10), with the help of which we write in simplified form:

$$(J_z)_{J_z J_z''} = J_z \delta_{J_z J_z''}, \qquad (J_z)_{J_z'' J_z'} = J_z' \delta_{J_z'' J_z'}$$

Substituting this into (30.17) and collecting all the terms of the equation in the left-hand side, we obtain

$$(J_x \pm iJ_y)_{J_z J_z'}(J_z' - J_z \pm 1) = 0 \quad (30.18)$$

Since one of the two factors in the left-hand side of this equation must be zero, we see that $J_x + iJ_y$ possesses only such matrix elements for which the column number is one less than the row number ($\hat{J}_z' = J_z - 1$), while $J_x - iJ_y$ possess only such matrix elements for which the column number is one greater than the row number ($\hat{J}_z' = J_z + 1$). At the same time, the eigenvalues of J_z vary only by unity (expressed in terms of h!) and pass through numbers from $-J$ to J. But in that case there are two and only two possibilities: J is either an integer or a half-integer (1/2, 3/2, etc.), since only integers or half-integers can, when reduced by an integral number of units, transform into themselves with the reverse sign.

Knowing which matrix elements $J_x \pm iJ_y$ are nonzero, let us now develop the matrix elements of $(J_x - iJ_y)(J_x + iJ_y)$. We begin by expanding the operator products:

$$(\hat{J}_x - i\hat{J}_y)(\hat{J}_x + i\hat{J}_y) = \hat{J}_x^2 + \hat{J}_y^2 + i(\hat{J}_x \hat{J}_y - \hat{J}_y \hat{J}_x)$$

$$= \hat{J}_x^2 + \hat{J}_y^2 - \hat{J}_z = \hat{J}^2 - \hat{J}_z^2 - \hat{J}_z \quad (30.19)$$

Since we have a diagonal matrix on the right, the matrix on the left is also diagonal. We write the diagonal element corresponding to this matrix,

$$(J_x - iJ_y)_{J_z, J_z+1}(J_x + iJ_y)_{J_z+1, J_z} = J^2 - J_z^2 - J_z \quad (30.20)$$

and apply Eq. (30.20) to the case when J_z is equal to its maximum value J. Then in the left-hand side both matrix elements involved in the product contain the indices $J + 1$. Hence, they are both equal

to zero, and we obtain the equation defining the eigenvalue of the angular momentum square:

$$J^2 = J (J + 1) \tag{30.21}$$

We should not be misled by the symbolic form of this equation, for we agreed to treat J^2 as the eigenvalue of the operator $\hat{J^2}$, and J as the maximum projection of the angular momentum on the z axis.

Unlike orbital angular momentum, total angular momentum can have both integral and half-integral values of J. Both are compatible with the commutation relations (30.8)[7].

Then it turns out that the angular momentum of a certain system can assume only integral or half-integral values, because J_z varies each time only by unity.

Let us now find the matrix elements of the angular momentum components. For this we note that

$$(J_x - iJ_y)_{J_z\,J_z+1} = (J_x + iJ_y)^{*}_{J_z+1\,J_z}$$

because $\hat{J}_x$ and $\hat{J}_y$ are Hermitian operators.

With the help of (30.20), this yields

$$| J_x + iJ_y |^2_{J_z+1\,J_z} = J (J + 1) - J_z (J_z + 1)$$
$$= (J - J_z) (J + J_z + 1) \tag{30.22a}$$

This matrix element becomes zero at $J_z = -(J + 1)$, as it should. Similarly

$$| J_x - iJ_y |^2_{J_z\,J_z+1} = J (J + 1) - J_z (J_z - 1)$$
$$= (J + J_z) (J - J_z + 1) \tag{30.22b}$$

Here we obtain zero when $J_z = J + 1$.

---

[7] Let us trace in greater detail how half-integral eigenvalues are eliminated in the case of orbital angular momentum. Let $k = l$, which corresponds to $J_z = J$. To operator $\hat{J}_x + i\hat{J}_y$ there corresponds $e^{i\varphi} \left(\dfrac{\partial}{\partial \vartheta} - \cot \vartheta \, \dfrac{1}{i} \dfrac{\partial}{\partial \varphi} \right) \equiv L_+$ (see Exercise 5, Section 24). Operating on $Y_l^l = P_l^l (\cos \vartheta) \, e^{il\varphi}$, we find that this operator should yield zero, as otherwise it would have a matrix element with indices l, $l + 1$, which must be equal to zero. Hence $\left(\dfrac{\partial}{\partial \vartheta} - l \cot \vartheta \right) P_l^l = 0$ and $P_l^l = C (\sin \vartheta)^l$. If, for example, $l = 1/2$, then $Y_{1/2}^{1/2} = (\sin \vartheta)^{1/2} \, e^{i\varphi/2}$. Operating on this wave function, we see that the operator $e^{i\varphi} \left(\dfrac{\partial}{\partial \vartheta} + \cot \vartheta \, \dfrac{1}{i} \dfrac{\partial}{\partial \varphi} \right) \equiv$ $\equiv L_-$ should yield the function $Y_{1/2}^{-1/2}$, since at $l = 1/2$ it has only this matrix element not equal to zero. But $L_- Y_{1/2}^{1/2} \neq Y_{1/2}^{-1/2}$, so that the value $l = 1/2$ is precluded for the orbital angular momentum.

If we define $(J_x + iJ_y)$ as a matrix with real elements, then the matrix $(J_x - iJ_y)$ must also have real elements: on this condition the commutation relations between the two matrices will correspond to (30.8). Therefore

$$(J_x + iJ_y)_{J_z \, J_z+1} = [(J - J_z)(J + J_z + 1)]^{1/2} \qquad (30.23a)$$

$$(J_x + iJ_y)_{J_z J_z+1} = [(J + J_z)(J - J_z + 1)]^{1/2} \qquad (30.23b)$$

Matrix $(J_x + iJ_y)$ has elements other than zero only next to the principal diagonal, to the right of it, while matrix $(J_x - iJ_y)$ has them only to the left of the principal diagonal, and also next to it. We shall require these matrices for $J = 1/2$. For this value of the total angular momentum its projection J_z takes only the values $1/2$ and $-1/2$. Accordingly, we obtain the following matrix elements:

$$(J_x + iJ_y)_{1/2 \, -1/2} = \left[\left(\frac{1}{2} + \frac{1}{2} \right) \left(\frac{1}{2} - \frac{1}{2} + 1 \right) \right]^{1/2} = 1$$
$$(30.24a)$$

$$(J_x - iJ_y)_{-1/2 \, 1/2} = \left[\left(\frac{1}{2} + \frac{1}{2} \right) \left(\frac{1}{2} - \frac{1}{2} + 1 \right) \right]^{1/2} = 1$$
$$(30.24b)$$

Each matrix has two rows and two columns.

From these matrices it is not difficult to find the matrices of the projections J_x and J_y. They are

$$J_x = \frac{1}{2} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \qquad (30.25a)$$

$$J_y = \frac{1}{2} \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \qquad (30.25b)$$

Multiplying J_y by $\pm i$ and adding with J_x, we return to (30.24a) and (30.24b), which justifies Eqs. (30.25a) and (30.25b).

We must supplement the matrices of J_x and J_y with that of J_z. Since it is by definition diagonal, it should be written in the form

$$J_z = \frac{1}{2} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \qquad (30.25c)$$

The Spin Variable of an Electron. We must now develop the state vector, or the wave function, on which the operators (30.25a)-(30.25c) will operate. It is apparent that these wave functions depend not on the spatial coordinates but on a special variable describing the spin degree of freedom. But since the matrices corresponding to angular momentum $J = 1/2$ have only two rows and two columns, the corresponding state vector has only two components. Let us call them ψ_1 and ψ_2.

Let us find the eigenfunctions corresponding to the projections of spin $\pm 1/2$. For this it is convenient to write them in a column:

$$\psi = \begin{pmatrix} \psi_1 \\ \psi_2 \end{pmatrix} \qquad (30.26)$$

These functions, taken together and written in Eq. (30.26) in the form of a single function ψ, must be eigenfunctions of the operator $\hat{J}_z$, defined as the matrix (30.25c):

$$\hat{J}_z \psi = J_z \psi, \qquad \frac{1}{2} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} \psi_1 \\ \psi_2 \end{pmatrix} = J_z \begin{pmatrix} \psi_1 \\ \psi_2 \end{pmatrix} \qquad (30.27)$$

Since $\hat{J}_z$ is a diagonal operator, the form of the eigenfunctions is easily determined:

$$J_z = 1/2, \qquad \psi_{1/2} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}; \qquad J_z = -1/2, \qquad \psi_{-1/2} = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \qquad (30.28)$$

that is

$$\psi_{1/2\ 1} = 1, \qquad \psi_{1/2\ 2} = 0, \qquad \psi_{-1/2\ 1} = 0, \qquad \psi_{-1/2\ 2} = 1$$

We thus arrive at a notation for the eigenfunctions of $\hat{J}_z$ which is quite similar to the general notation for the eigenfunctions of any operator, for example the operator $\hat{\lambda}$ in terms of x: $\psi(\lambda, x)$.

For the independent variable x we have the indices 1 and 2, and for the spin operator eigenvalues, $1/2$ and $-1/2$. As distinct from the variable x, which assumes a continuous set of values, the spin variable has only two values (1 and 2), but nevertheless all the formal demands imposed on operators and eigenvalues in quantum mechanics are satisfied in the case of spin. For example, it can be seen from the notation of the operators $\hat{J}_x$, $\hat{J}_y$, and $\hat{J}_z$ that they are Hermitian. The wave functions corresponding to different eigenvalues of the operator $\hat{J}_z$ are orthogonal. This can be seen directly from (30.28): if we denote the spin variable s ($s = 1, 2$) and the eigenvalues J_z ($J_z = 1/2, -1/2$), the orthonormality relations

$$\sum_{s=1}^{2} \psi^*_{J_z\ s} \psi_{J'_z\ s} = \delta_{J_z J'_z} \qquad (30.29)$$

are satisfied.

Thus, the eigenvalue of the spin projection should be seen as a fourth quantum number in addition to n, l, and k. In order to avoid writing the fraction $1/2$ every time, usually only the sign of the spin projection, denoted by the letter σ, is stated. In normalizing an arbitrary wave function, the square of its modulus is integrated over the spatial variables and summed over the spin variable s. For example, the orthonormality condition in a central field should be

written as

$$\sum_s \int \psi^*_{n'l'k'\sigma'} \psi_{nlk\sigma} \, dV = \delta_{nn'} \delta_{ll'} \delta_{kk'} \delta_{\sigma\sigma'} \tag{30.30}$$

Pauli Spin Matrices. Having agreed not to write the magnitude of the spin projection, only its sign, it is convenient to introduce in place of J_x, J_y, and J_z three such matrices:

$$\sigma_x \equiv \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \qquad \sigma_y \equiv \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \qquad \sigma_z \equiv \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \tag{30.31}$$

which were introduced by Pauli for the quantum mechanical description of electron spin. Let us examine the properties of these matrices more closely.

Note, first of all, that each of them has one nonzero element in each row and in each column. This makes it possible to represent the operation of such matrices on the ψ function (30.26) in the form of a substitution:

$$\hat{\sigma}_x \psi = \hat{\sigma}_x \begin{pmatrix} \psi_1 \\ \psi_2 \end{pmatrix} = \begin{pmatrix} \psi_2 \\ \psi_1 \end{pmatrix}$$

$$\hat{\sigma}_y \psi = \begin{pmatrix} -i\psi_2 \\ i\psi_1 \end{pmatrix}, \qquad \hat{\sigma}_z \psi = \begin{pmatrix} \psi_1 \\ -\psi_2 \end{pmatrix} \tag{30.32}$$

It is easier to find the rules for multiplying the matrices with the help of substitutions than in the general way; for example

$$\hat{\sigma}_x^2 \psi = \hat{\sigma}_x \begin{pmatrix} \psi_2 \\ \psi_1 \end{pmatrix} = \begin{pmatrix} \psi_1 \\ \psi_2 \end{pmatrix} = \psi \tag{30.33a}$$

Obviously, this equation should be symbolically written as $\hat{\sigma}_x^2 = 1$. The unity here denotes $\delta_{ss'}$. In the same way we find that

$$\hat{\sigma}_y^2 \psi = \hat{\sigma}_y \begin{pmatrix} -i\psi_2 \\ i\psi_1 \end{pmatrix} = \begin{pmatrix} (-i)\,i\psi_1 \\ i\,(-i)\,\psi_2 \end{pmatrix} = \begin{pmatrix} \psi_1 \\ \psi_2 \end{pmatrix} = \psi \tag{30.33b}$$

and finally

$$\hat{\sigma}_z^2 \psi = \hat{\sigma}_z \begin{pmatrix} \psi_1 \\ -\psi_2 \end{pmatrix} = \begin{pmatrix} \psi_1 \\ \psi_2 \end{pmatrix} = \psi \tag{30.33c}$$

Now, in the same way we find the paired products of the Pauli matrices:

$$\hat{\sigma}_x\hat{\sigma}_y\psi = \hat{\sigma}_x\begin{pmatrix} -i\psi_2 \\ i\psi_1 \end{pmatrix} = \begin{pmatrix} i\psi_1 \\ -i\psi_2 \end{pmatrix} = i\begin{pmatrix} \psi_1 \\ -\psi_2 \end{pmatrix}$$

$$= i\hat{\sigma}_z\psi = -\hat{\sigma}_y\hat{\sigma}_x\psi \tag{30.34a}$$

$$\hat{\sigma}_z\hat{\sigma}_x\psi = \hat{\sigma}_z\begin{pmatrix} \psi_2 \\ \psi_1 \end{pmatrix} = \begin{pmatrix} \psi_2 \\ -\psi_1 \end{pmatrix} = i\begin{pmatrix} -i\psi_2 \\ i\psi_1 \end{pmatrix}$$

$$= i\hat{\sigma}_y\psi = -\hat{\sigma}_x\hat{\sigma}_z\psi \tag{30.34b}$$

$$\hat{\sigma}_y\hat{\sigma}_z\psi = \hat{\sigma}_y\begin{pmatrix} \psi_1 \\ -\psi_2 \end{pmatrix} = \begin{pmatrix} i\psi_2 \\ i\psi_1 \end{pmatrix} = i\begin{pmatrix} \psi_2 \\ \psi_1 \end{pmatrix}$$

$$= i\hat{\sigma}_x\psi = -\hat{\sigma}_z\hat{\sigma}_y\psi \tag{30.34c}$$

All the obtained equations can be written as one equation in tensor form (see Sec. 11):

$$\hat{\sigma}_l\hat{\sigma}_m = \delta_{lm} + \varepsilon_{lmn}i\hat{\sigma}_n \tag{30.35}$$

This means that any expression that is quadratic with respect to the Pauli matrices can be reduced to a linear expression. For example

$$(\mathbf{A}\cdot\hat{\sigma})(\mathbf{B}\cdot\hat{\sigma}) = A_l\hat{\sigma}_l \times B_m\hat{\sigma}_m = A_lB_l + i\varepsilon_{lmn}A_lB_m\hat{\sigma}_n$$

$$= (\mathbf{A}\cdot\mathbf{B}) + i\,(\mathbf{A}\times\mathbf{B})\,\hat{\sigma} \tag{30.36}$$

Of course, here both Pauli matrices operate on the spin variables of the same electron.

The Vector Properties of Pauli Matrices. We shall show that the Pauli matrices can be treated as vector components. For this we must verify that in rotations of the coordinate system they transform like vector components, that is, that the transformed operators possess the same properties as the initial ones.

In Section 9 we introduced symbols for the cosines of the angles between the old and new coordinate axes, namely, the cosine between the old axis α and the new axis α' was denoted (α', α). Then the components of the transformed vector are expressed in terms of the vector components with respect to the old axes in the following way:

$$\hat{\sigma}'_{\alpha'} = (\alpha', \alpha)\,\hat{\sigma}_\alpha, \qquad \hat{\sigma}'_{\beta'} = (\beta', \beta)\,\hat{\sigma}_\beta \tag{30.37}$$

We must prove that their product yields a result similar to (30.35).

Multiplying $\hat{\sigma}'_{\alpha'}$ by $\hat{\sigma}'_{\beta'}$, we obtain

$$\hat{\sigma}'_{\alpha'}\hat{\sigma}'_{\beta'} = (\alpha', \alpha)(\beta', \beta)\,\hat{\sigma}_\alpha\hat{\sigma}_\beta$$

$$= (\alpha', \alpha)(\beta', \beta)(\delta_{\alpha\beta} + i\varepsilon_{\alpha\beta\gamma}\hat{\sigma}_\gamma)$$

$$= \delta_{\alpha'\beta'} + i(\alpha', \alpha)(\beta', \beta)\,\varepsilon_{\alpha\beta\gamma}\hat{\sigma}_\gamma$$

But $\varepsilon_{\alpha\beta\gamma}$ is an invariant tensor and is the same in any coordinate system. The same can be said if one of the tensor indices (γ) is contracted (summed over) with the index of some vector, for example, $\hat{\sigma}_\gamma$. This follows simply from the properties of the transformation coefficients (α', α), ..., (γ', γ). Therefore

$$(\alpha', \alpha)(\beta', \beta)\,\varepsilon_{\alpha\beta\gamma}\hat{\sigma}_\gamma = \varepsilon_{\alpha'\beta'\gamma'}\hat{\sigma}_{\gamma'}$$

Thus

$$\hat{\sigma}'_{\alpha'}\hat{\sigma}'_{\beta'} = \delta_{\alpha'\beta'} + i\varepsilon_{\alpha'\beta'\gamma'}\hat{\sigma}_{\gamma'}$$

the same as for the old components $\hat{\sigma}_\alpha$, $\hat{\sigma}_\beta$.

The transformation formulas for a rotation around one axis are conveniently written in explicit form:

$$\hat{\sigma}'_x = \hat{\sigma}_y \cos\varphi + \hat{\sigma}_y \sin\varphi$$
$$\hat{\sigma}'_y = -\hat{\sigma}_x \sin\varphi + \hat{\sigma}_y \cos\varphi \tag{30.38}$$

We shall use these formulas to find the transformation law for the wave function components ψ_1, ψ_2.

It was shown in Section 26 that the transformation of a rotation of coordinate axes belongs to the class of unitary transformations. Equations (30.37) and (30.38) express unitary transformations of operators. But it is of interest to obtain the corresponding unitary transformation of the wave functions themselves.

We begin with the simplest transformation: a rotation around the z axis, which for operators has the form (30.38). We shall proceed from the standard wave functions (30.28), which we shall denote simply ψ_+ and ψ_- so as to avoid writing fractions in the index. An arbitrary function of the spin variable can be represented as a linear combination of ψ_+ and ψ_- thus:

$$\psi = x_1\psi_+ + x_2\psi_- \tag{30.39}$$

Function ψ is, of course, also a two-component one, its first component, as can be seen from (30.28) and (30.39), being equal to x_1, and its second to x_2.

We now carry out a rotation of the coordinate system around the z axis through an infinitesimal angle $\delta\varphi$. Then, as we know from

the definition of angular momentum (30.2) or (30.3),

$$\delta\psi_+ = \frac{i}{2}\,\delta\varphi\,\psi_+, \qquad \delta\psi_- = -\frac{i}{2}\,\delta\varphi\,\psi_- \tag{30.40}$$

Hence, for $\delta\psi$ we have

$$\delta\psi = \frac{i}{2}\,\delta\varphi \begin{pmatrix} x_1 \\ -x_2 \end{pmatrix} \tag{30.41}$$

Integrating this relation, we obtain the transformation formula for ψ' for a finite rotation angle φ:

$$\psi' \equiv \begin{pmatrix} x_1' \\ x_2' \end{pmatrix} = \begin{pmatrix} x_1 e^{i\varphi/2} \\ x_2 e^{-i\varphi/2} \end{pmatrix} \tag{30.42}$$

Thus, the two-component function $\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$ transforms through one-half the rotation angle. Note that $|\psi'|^2 = |x_1'|^2 + |x_2'|^2 = |\psi|^2 = |x_1|^2 + |x_2|^2$, as should be in a unitary transformation retaining the normalization of the wave function.

Let us find the more general form of the unitary transformation of the wave function ψ, which we represent with the help of four numbers, α, β, γ, and δ:

$$x_1' = \alpha x_1 + \beta x_2$$
$$x_2' = \gamma x_1 + \delta x_2 \tag{30.43}$$

The complex conjugate quantities $x_1'^*$, $x_2'^*$ can be expressed respectively in terms of α^*, β^*, γ^*, and δ^*. Let us require that $|\psi'|^2 = |\psi|^2$. On the transformation coefficients this imposes the following four conditions, which appear from a comparison of the expressions multiplying $|x_1|^2$, $|x_2|^2$, $x_1 x_2^*$, and $x_1^* x_2$:

$$\alpha^*\alpha + \gamma^*\gamma = 1, \qquad \alpha^*\beta + \gamma^*\delta = 0$$
$$\beta^*\beta + \delta^*\delta = 1, \qquad \alpha\beta^* + \gamma\delta^* = 0 \tag{30.44}$$

From the conditions (30.44) there follows the relation between α, β, γ, δ and the complex conjugate quantities α^*, β^*, γ^*, δ^*:

$$\begin{pmatrix} \alpha^* & \beta^* \\ \gamma^* & \delta^* \end{pmatrix} = D^{-1} \begin{pmatrix} \delta & -\gamma \\ -\beta & \alpha \end{pmatrix} = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}^{-1} \tag{30.45}$$

where $D = \alpha\delta - \beta\gamma$, that is, the determinant of the transformation. Equation (30.45) denotes that the complex conjugate transformation matrix is equal to the inverse matrix, but this is precisely the condition for unitarity. Forming the determinants of both sides of (30.45), we find that $D^* = D^{-1}$. We can put $D = 1$ without restricting the generality; then

$$\alpha = \delta^*, \qquad \beta = -\gamma^* \tag{30.46}$$

Let us apply the obtained relationships to a rotation of a coordinate system through an angle ϑ about the x axis. Suppose the projection of the spin on the initial z axis was $+1/2$. The mean value of the spin projection on the new axis is equal to $\cos \vartheta$, since the relations between mean values in quantum mechanics are the same as between the values themselves in classical mechanics. Since the basic properties of operators are conserved in a rotation of the coordinate axes, we assume that the operator of the spin projection on the new axis z (σ'_z) has the form (30.31). Let us find the expressions for the wave functions transformed in the rotation.

If the wave function of the state was $\psi_+ = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ with respect to the old coordinate system, with respect to the new axes we should have, from (30.43),

$$\psi'_1 = \alpha, \qquad \psi'_2 = \gamma, \qquad \psi' = \begin{pmatrix} \alpha \\ \gamma \end{pmatrix}$$

and for the complex conjugate we obtain, with the help of (30.46),

$$\psi'^* = \begin{pmatrix} \alpha^* \\ \gamma^* \end{pmatrix} = \begin{pmatrix} \delta \\ -\beta \end{pmatrix}$$

Both these expressions should be substituted into the definition of the mean value of the angular momentum projection on the new z axis:

$$\langle \sigma'_z \rangle = \sum_s \psi'^*_s \hat{\sigma}_z \psi'_s = \cos \vartheta \tag{30.47}$$

Using the known form of the operator $\hat{\sigma}_z$ and the expressions for ψ'^* and ψ', we determine from (30.47)

$$\alpha\delta + \beta\gamma = \cos \vartheta$$

Since $\alpha\delta - \beta\gamma = D = 1$, we obtain the second equation relating the transformation coefficients: $\alpha\delta - \beta\gamma = 1$. Adding and subtracting these equations, we find

$$\alpha\delta = \frac{1}{2}(1 + \cos \vartheta) = \cos^2 \frac{\vartheta}{2}, \qquad \beta\gamma = -\sin^2 \frac{\vartheta}{2}$$

Applying (30.46), we obtain

$$\alpha^*\alpha = |\alpha|^2 = \cos^2 \frac{\vartheta}{2}, \qquad \beta^*\beta = |\beta|^2 = \sin^2 \frac{\vartheta}{2}$$

We assume that α is a real quantity equal to $\cos(\vartheta/2)$. Then, to satisfy the condition $\alpha\delta - \beta\gamma = 1$, we must consider β to be a purely imaginary quantity $i \sin(\vartheta/2)$. The correctness of this choice of α and β is seen from the following. We find $\langle \sigma_y \rangle$, which

evidently must be equal to sin ϑ. On the other hand we have

$$\langle \sigma_y \rangle = \sum_s \psi_s'^* \hat{\sigma}_y \psi_s = -i\psi_1'^* \psi_2 + i\psi_2'^* \psi_1$$

$$= 2 \sin \frac{\vartheta}{2} \cos \frac{\vartheta}{2} = \sin \vartheta \tag{30.48}$$

We have thus obtained a matrix describing the rotation about the x axis:

$$\begin{pmatrix} \cos \dfrac{\vartheta}{2} & i \sin \dfrac{\vartheta}{2} \\ i \sin \dfrac{\vartheta}{2} & \cos \dfrac{\vartheta}{2} \end{pmatrix}$$

The most general form of the matrix is obtained by performing one more rotation about the new axis z' through an angle χ. The three rotation angles φ, ϑ, and χ are essentially the Euler angles (Sec. 9). Multiplying all three matrices obtained for each rotation separately, we obtain the most general rotation matrix:

$$\begin{pmatrix} e^{i\chi/2} & 0 \\ 0 & e^{-i\chi/2} \end{pmatrix} \times \begin{pmatrix} \cos \dfrac{\vartheta}{2} & i \sin \dfrac{\vartheta}{2} \\ i \sin \dfrac{\vartheta}{2} & \cos \dfrac{\vartheta}{2} \end{pmatrix} \times \begin{pmatrix} e^{i\varphi/2} & 0 \\ 0 & e^{-i\varphi/2} \end{pmatrix}$$

$$= \begin{pmatrix} \cos \dfrac{\vartheta}{2} e^{i(\varphi+\chi)/2} & i \sin \dfrac{\vartheta}{2} e^{i(\chi-\varphi)/2} \\ i \sin \dfrac{\vartheta}{2} e^{i(\varphi-\chi)/2} & \cos \dfrac{\vartheta}{2} e^{-i(\varphi+\chi)/2} \end{pmatrix} \tag{30.49}$$

Such a matrix has a certain analogy with the matrix of cosines (λ, μ). But unlike the latter, it operates not on the components of a vector but on the components of a special quantity, called a *spinor*. A spinor is a two-component complex quantity with a transformation law described by matrix (30.49), whereas a vector is a three-component quantity which transforms in rotations of the coordinates with the help of a matrix of cosines.

Since the spinor transformation matrix is expressed in terms of half-angles, the spinor does not revert to its initial value in a rotation of the coordinate axis through 360°. It is clear from this that no classical directly measurable quantity can be expressed linearly in terms of a spinor. Only bilinear expressions of the type (30.47) and (30.49) are possible. Such quantities, naturally, assume their initial values in a rotation through 360°. But only they can be measured in quantum mechanics, and by no means the wave functions themselves.

The Operator of the Total Angular Momentum of an Electron. We examined the spin of an electron apart from its orbital angular momentum. Let us now form the vector operator of the total angular

momentum of an electron:

$$\hat{\mathbf{j}} = \hat{\mathbf{M}} + \frac{1}{2}\,\hat{\boldsymbol{\sigma}} \tag{30.50a}$$

or in terms of the components

$$\hat{\mathbf{j}}_x = \hat{M}_x + \frac{1}{2}\,\hat{\sigma}_x, \quad \hat{\mathbf{j}}_y = \hat{M}_y + \frac{1}{2}\,\hat{\sigma}_y, \quad \hat{\mathbf{j}}_z = \hat{M}_z + \frac{1}{2}\,\hat{\sigma}_z \tag{30.50b}$$

Operators $\hat{\mathbf{M}}$ and $\hat{\boldsymbol{\sigma}}$ commute, since they operate on different variables: $\hat{\mathbf{M}}$ on a spatial variable, and $\hat{\boldsymbol{\sigma}}$ on a spin variable. Since $\hat{M}_i$ and $\hat{\sigma}_i$ have the same commutation relations, the components of the vector operator $\hat{\mathbf{j}}$ have the same commutation relations. Furthermore, $\hat{\mathbf{j}}$ is a vector, because $\hat{\mathbf{M}}$ and $\hat{\boldsymbol{\sigma}}$ transform in the same way in the rotations of the coordinate system.

If the eigenvalue of the square of the orbital angular momentum is not zero, that is, if $l \neq 0$, the eigenvalues of the square of the total angular momentum may be equal to either $(l + 1/2)(l + 3/2)$ or $(l - 1/2)(l + 1/2)$. In the former case the spin and the orbital angular momentum are said to be parallel, in the latter case they are antiparallel.

This leads to a doubling of the number of electron states. Instead of the quantum number σ we can use the quantum number $j = l \pm 1/2$.

Spin Magnetic Moment. The spin of an electron, like its orbital angular momentum, is associated with a definite magnetic moment. But experiment shows that the ratio of spin magnetic moment to spin angular momentum is twice as great as for the electron's orbital motion. There is nothing paradoxical in this because the result (17.30) can be applied only to orbital angular momentum. At the same time we can deduce the spin magnetic moment from the Dirac relativistic wave equation for an electron (Sec. 37); in agreement with experiment, the relation between the spin magnetic moment and the spin angular momentum is

$$\boldsymbol{\mu}_\sigma = \frac{e}{2mc}\,\boldsymbol{\sigma} \tag{30.51}$$

Hence, the projection of spin magnetic moment on any axis is

$$(\boldsymbol{\mu}_\sigma)_z = \pm\,\frac{eh}{2mc} \equiv \pm\,\beta \tag{30.52}$$

(Here we again measure the angular momentum in conventional units.) The quantity β is called the *Bohr magneton*. It is a natural unit of magnetic moment.

Since the spin and orbital moments produce their own magnetic moments, magnetic interaction occurs between them. It is proportional to the product of the magnetic moments. But the formula

of each magnetic moment involves the velocity of light in the denominator. Hence, the *spin-orbit coupling*, as it is called, is inversely proportional to c^2, that is, it is a relativistic effect. If the velocities of atomic electrons are small in comparison with c, their magnetic interaction is small. This is always the case for light atoms.

Due to magnetic forces, the energy level with $j = l + 1/2$ always differs slightly from the energy level with $j = l - 1/2$. In such simple form this conclusion refers to a separate electron in a central field, for example, in an atom of an alkali metal.

EXERCISES

1. Show that the common general property of the Pauli spin matrices, $\hat{\sigma}_\alpha \hat{\sigma}_\beta + \hat{\sigma}_\beta \hat{\sigma}_\alpha = 2\delta_{\alpha\beta}$, is conserved in rotations of the coordinate system, making use of the fact that

$$(\alpha, \nu)(\beta, \nu) = \delta_{\alpha\beta}$$

2. Determine the eigenvalues of the scalar product $(\hat{\boldsymbol{\sigma}}_1 \cdot \hat{\boldsymbol{\sigma}}_2)$ for two electrons, making use of the fact that $\hat{\boldsymbol{\sigma}}_1$ and $\hat{\boldsymbol{\sigma}}_2$ commute.

Solution. Since $\hat{\boldsymbol{\sigma}}_1$ and $\hat{\boldsymbol{\sigma}}_2$ commute, the conventional formula

$$(\hat{\boldsymbol{\sigma}}_1 + \hat{\boldsymbol{\sigma}}_2)^2 = \sigma_1^2 + \sigma_2^2 + 2(\hat{\boldsymbol{\sigma}}_1 \cdot \hat{\boldsymbol{\sigma}}_2)$$

holds true.

We know from Section 29 that the sum of two angular momenta varies from their sum to their difference, so that the total spin assumes values 1 and 0 (according to the largest projection). Since $\hat{\boldsymbol{\sigma}}$ is the double operator of spin, the respective maximum projections are all twice as large. Hence, when the spins are added like parallel vectors, the eigenvalue of the square, $(\hat{\boldsymbol{\sigma}}_1 + \hat{\boldsymbol{\sigma}}_2)^2$, is fourfold the corresponding square of the angular momentum, that is, it is equal to $4 \times 1 \times 2 = 8$; when the spins are antiparallel it is zero. Furthermore, $\hat{\sigma}_x^2 = \hat{\sigma}_y^2 = \hat{\sigma}_z^2 = 1$, so that the eigenvalue of $(\hat{\boldsymbol{\sigma}}_1 \cdot \hat{\boldsymbol{\sigma}}_2)$ is $(8 - 6)/2 = 1$ for parallel spins, and $(0 - 6)/2 = -3$ for antiparallel spins.

3. Determine the eigenfunctions of the operators $\hat{\sigma}_x$ and $\hat{\sigma}_y$.

Answer.

$$\sigma_x = \frac{1}{2}, \quad \psi = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ 1 \end{pmatrix}; \quad \sigma_x = -\frac{1}{2}, \quad \psi = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ -1 \end{pmatrix};$$

$$\sigma_y = \frac{1}{2}, \quad \psi = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ i \end{pmatrix}$$

$$\sigma_y = -\frac{1}{2}, \quad \psi = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ -i \end{pmatrix}$$

Thus, the eigenfunctions of all three noncommutative operators are different.

31

THE QUASI-CLASSICAL APPROXIMATION

Application of quantum mechanics to specific problems frequently encounters formidable mathematical difficulties. In such cases it may prove necessary to make use of various approximate methods. Such methods should not be seen simply as a forced substitution of exact solutions. On the contrary, an analysis of the approximation makes it possible to gain a deeper insight into the main aspects of the problem and distinguish between the main and the secondary. When the approximation is found to be lacking we usually find that the initial simplifying assumptions were not valid.

In this section we shall examine an approximate method of quantum mechanics, which is applicable when the problem in hand has a close classical correspondence.

The Quasi-Classical Approximation. It was shown in Section 23 that the limiting transition from quantum to classical mechanics is achieved by the substitution

$$\psi = e^{iS/h} \tag{31.1}$$

where S is the action of a particle. In order to perform the limiting transition we must formally assume $h = 0$. Suppose, now, that the action quantum is small, but finite, in comparison with the characteristic quantities of the action dimensions in the problem in hand. Then S cannot be considered strictly equal to the classical action and should be represented as an expansion:

$$S = S_0 + h S_1 + h^2 S_2 + \ldots \tag{31.2}$$

Here, S_0 is the classical action, while all the other terms are quantum corrections. In specific problems we usually have to deal, in addition to S_0, with only the first term of the expansion S_1, because if other terms are of the same order as the first and second terms, then the approximation itself is not justified.

Substituting the expansion (31.2) into Eq. (31.1), and then into the Schrödinger one-dimensional equation, we find S_0, S_1, S_2, ... for motion with one degree of freedom. In the case when the variables separate in the Schrödinger equation for a system with several degrees of freedom, as, for example, in a central field, substitutions similar to (31.1) and (31.2) can be performed for each degree of freedom separately.

After cancelling out $\psi = e^{iS/h}$, we obtain from the Schrödinger equation the exact equation for S, as yet without approximations:

$$(S')^2 + \frac{h}{i} S'' = 2m (E - U) \equiv p^2 \qquad (31.3)$$

We substitute the series (31.2) into this equation and collect the terms multiplying the same powers of h to get

$$(S_0')^2 - p^2 + h \left(2S_0'S_1' + \frac{1}{i} S_0'' \right)$$
$$+ h^2 \left(2S_0'S_2' + (S_1')^2 + \frac{1}{i} S_1'' \right) + \ldots = 0 \qquad (31.4)$$

Equating the factors multiplying successive powers of h to zero, we find the equations for S_0, S_1, S_2, ..., in which the classical momentum $p = [2m (E - U)]^{1/2}$ is taken as the known quantity. Note that in each new approximation there appears a corresponding new function, so that in principle they may all be defined successively:

$$S_0' = \pm p \qquad (31.5)$$

$$S_1' = \frac{i}{2} \frac{S_0''}{S_0'} = \frac{i}{2} (\ln S_0')' = i \left(\ln \sqrt{S_0'} \right)' \qquad (31.6)$$

$$S_2' = \left| \pm \frac{1}{p^{1/2}} \left\{ [(\ln \sqrt{p})']^2 - (\ln \sqrt{p})'' \right\} \right. \qquad (31.7)$$

It can be seen from the latter equation that the approximation becomes invalid when the classical momentum vanishes, that is, at the *turning points*[8] $E = U (x)$. Then the third term of the expansion becomes infinite. Furthermore, all the expansion terms become large when the derivatives p', p'', ..., expressed in terms of the potential energy derivatives U', U'', ..., are large. In other words, for an approximation to be valid, the force $F = -U'$ and its coordinate-derivatives should not be large. Expansion (31.2) can be used for a sufficiently smooth potential energy curve. This assumption is usually satisfied in real problems. But almost always one encounters the turning points in whose neighbourhood additional investigation is required.

We first suppose that a solution is found for such x's that are sufficiently far away from the turning points. Then, retaining the first and second terms in (31.2), we obtain

$$\psi^{\pm} = \exp \left(\pm \frac{i}{h} \int p \, dx - \ln \sqrt{p} \right)$$
$$= \frac{1}{p^{1/2}} \exp \left(\pm \frac{i}{h} \int p \, dx \right) \qquad (31.8)$$

---

[8] At these points a particle moving according to classical laws should have to alter the direction of motion, or turn.

This solution coincides, up to the factor $p^{-1/2}$ with the conventional representation of the wave function ψ in terms of the action:

$\psi = e^{iS/h}$, $S = \int p \, dx$. The two signs in the formula correspond

to waves travelling in both directions. The approximation (31.8) is termed quasi-classical.

The Solution in the Classically Inaccessible Domain. Strictly speaking, however, the term can be applied only to those spatial domains where the momentum p is a real quantity. In those regions
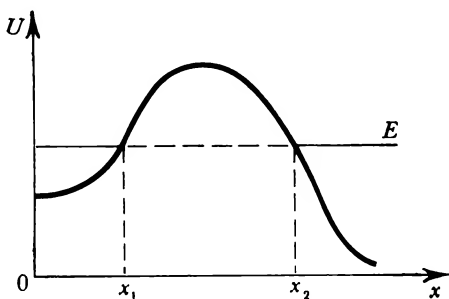
Figure 35

where $U > E$ the momentum becomes imaginary, and all that remains is a purely superficial similarity between quantum and classical formulas. In the problem on a potential well of finite depth (Sec. 28) it was shown that there is a finite probability of a particle occurring in a classically inaccessible domain, where $U > E$. For the case of a rectangular well, the solution is achieved with the help of boundary conditions, making it possible to match the wave function for domains where $E > U$ and $U > E$. We shall now consider a typical problem in which it is necessary to match the wave functions of a quasi-classical approximation between the domains where $p^2 > 0$ and where $p^2 < 0$.

Figure 35 presents the potential energy when the wave function is of special interest in the classically inaccessible domain. The total energy of the particle is less than the maximum value of the potential energy. The wave function of a particle located to the left of the hump cannot decrease to zero along the finite distance from point x_1 to point x_2. Consequently, neither is the wave function zero to the right of point x_2, where the total energy is again greater than the potential energy. Here, the wave function ceases to fall off at all, so that the motion is on the whole infinite. This means that a potential barrier of finite width cannot retain a particle in a well infinitely.

Let us show how to calculate the transmission coefficient of a potential barrier, that is, the probability of a particle passing through it.

The Wave Function in the Neighbourhood of the Turning Points. In matching the wave function in the regions where $p^2 > 0$ and where $p^2 < 0$ we encounter the following difficulty: the solution (31.8) is not valid in the neighbourhood of the two turning points (x_1 and x_2). It is valid only at a sufficiently great distance from these points, on either side of both. We can, however, assume that up to some points, distant from x_1 and x_2, the potential energy curve is approximated by its tangents through the turning points though, of course, the domains of such linear dependence of U on x are still very small in comparison with the whole domain below the barrier. We write the stated expansion of the potential energy as follows:

$$U(x) = U(x - x_1 + x_1) = U(x_1) + (x - x_1)\left(\frac{dU}{dx}\right)_1$$

$$= U(x_1) - (x - x_1)F_1 \qquad (31.9)$$

where F_1 is the force acting on the particle at point x_1, $U(x_1) = E$. We substitute the expansion (31.9) into the Schrödinger equation to get

$$-\frac{h^2}{2m}\frac{d^2\psi}{dx^2} = (E - U)\psi = (x - x_1)F_1\psi \qquad (31.10)$$

At a sufficient distance from the turning point we can make use of the quasi-classical approximation (31.8) and represent the solution in the form

$$\psi^\pm = \frac{1}{[(x - x_1)F_1]^{1/4}}\exp\left(\pm\frac{i}{h}\int_{x_1}^{x}[2m(x - x_1)F_1]^{1/2}\,dx\right)$$

$$= \frac{1}{[(x - x_1)F_1]^{1/4}}\exp\left(\pm\frac{i}{h}\times\frac{2}{3}(2mF_1)^{1/2}(x - x_1)^{3/2}\right) \qquad (31.11)$$

Although the integral involves the lower limit x_1, close to $x = x_1$ the solution (31.11) is not valid. At the turning point $x = x_1$ in Figure 35, the derivative $(dU/dx)_1$ is positive. Here, the force of attraction $F_1 = -(\partial U/\partial x)_1 < 0$, and therefore to the left of point x_1 the radicand is positive. To the right of the turning point the radicand is imaginary. Here the solution has the form

$$\psi^\pm = \frac{1}{[(x - x_1)F_1]^{1/4}}\exp\left(\mp\frac{2}{3h}(2m|F_1|)^{1/2}(x - x_1)^{3/2}\right) \qquad (31.12)$$

Both solutions (31.11) and (31.12) are valid only where the quantities in the exponents are sufficiently large in magnitude in comparison with unity. But from (31.11) and (31.12) it is not apparent how the solution passes through the intermediate domain, that is, the form a solution far to the left of x_1 (say, ψ^+) assumes when x varies to values far to the right of x_1.

In order to link together two asymptotic solutions we have to know the exact solution for the intermediate domain and extend it to domains lying far away from the turning point. If on the left the exact solution takes, say, the form ψ^+, on the right we obtain a linear combination of ψ^+ and ψ^-. But these are asymptotic forms of one and the same solution, and they should consequently be used to match the wave functions on both sides of the turning point.

Let us now show how the exact solution is developed. Equation (31.10) always allows for an exact solution, since it involves the independent variable linearly. Before we proceed to find the solution, let us make the following substitution:

$$\xi = \left(\frac{2mF_1}{\hbar^2} \right)^{1/3} (x - x_1) \tag{31.13}$$

so as to get rid of extra letters cluttering up the equation. It can then be rewritten as follows:

$$\frac{d^2\psi}{d\xi^2} + \xi\psi = 0 \tag{31.14}$$

The solution is conveniently sought in the form

$$\psi = \int e^{iq\xi} f(q)\, dq \tag{31.15}$$

where the integration limits are not stated for the time being. We substitute (31.15) into (31.14), performing the following operations:

$$\psi'' = - \int e^{iq\xi} q^2 f(q)\, dq$$

$$\xi\psi = \int \xi e^{iq\xi} f(q)\, dq = - i \int \frac{\partial}{\partial q} (e^{iq\xi})\, f(q)\, dq$$

$$= - i e^{iq\xi} f(q) \Big| + i \int e^{iq\xi} \frac{df}{dq}\, dq$$

We select the integration limits such that the integrated expression vanishes. Then substitution into the differential equation (31.14) yields

$$\int \left(i \frac{df}{dq} - q^2 f \right) e^{iq\xi}\, dq = 0 \tag{31.16}$$

Since this equation must hold at all values of ξ, the function f satisfies the equation

$$i\,\frac{df}{dq} - q^2 f = 0 \tag{31.17}$$

so that

$$f = \exp\left(i\int q^2\,dq\right) = e^{iq^3/3} \tag{31.18}$$

whence

$$\psi = \int e^{i(q\xi + q^3/3)}\,dq \tag{31.19}$$

This solution holds for any ξ's, positive as well as negative. By carrying out certain computations based on the theory of functions of a complex variable, we can show that for large absolute values of $|\xi|$ the obtained solution passes into a solution of the form (31.11) and (31.12), between which there is the following correspondence:

solution to the left of x_1: | solution to the right of x_1:

$$\frac{1}{\xi^{1/4}}\sin\left(\frac{2}{3}\,\xi^{3/2} + \frac{\pi}{4}\right) \to \frac{1}{2\,|\xi|^{1/4}}\,e^{-2|\xi|^{3/2}/3} \tag{31.20a}$$

$$\frac{1}{\xi^{1/4}}\cos\left(\frac{2}{3}\,\xi^{3/2} + \frac{\pi}{4}\right) \to \frac{1}{2\,|\xi|^{1/4}}\,e^{2|\xi|^{3/2}/3} \tag{31.20b}$$

Substituting x for ξ, we see that the solutions on the left are formed by linear combinations of ψ^+, ψ^- expressed according to (31.11). In turn, the exponents in (31.11) at even greater distances from the turning points of rotation should be replaced by the integrals $\int p\,dx$ and $\int |p|\,dx$.

Penetration of the Potential Barrier. In Exercise 2, Section 28, we obtained the exact formula for the probability of a particle tunneling a rectangular potential barrier. We shall now find the formula for the probability of tunneling a potential barrier of arbitrary configuration in the quasi-classical approximation.

The possibility of passing potential barriers is a general property of motion in quantum mechanics, associated with the fact that the wave function does not vanish in the classically inaccessible region, where $E < U$. In the most general case, to determine the probability of passing below the barrier, it is necessary to solve the Schrödinger equation for the given problem. Accordingly, a general exact probability formula cannot be developed. In the quasi-classical approximation, however, a general formula can be obtained. We shall now proceed with its deduction.

Since all the exponents in this approximation are large in comparison with unity, we may conclude that the probability of penetrating

the barrier must be exponentially small, otherwise the very approximation is inapplicable. But if the probability of penetrating the barrier is very small, the probability of reflection from it is close to unity. Hence, to the left of the barrier, from where the particles approach it, the wave function can, to a high degree of accuracy, be replaced by a standing wave of the form (28.5):

$$p^{1/2}\psi = \exp\left(-\frac{i}{h}\int_{x_1}^{x} p\,dx - \frac{i\pi}{4}\right) + \exp\left(\frac{i}{h}\int_{x_1}^{x} p\,dx + \frac{i\pi}{4}\right)$$

$$= 2\sin\left(\frac{1}{h}\int_{x}^{x_1} p\,dx + \frac{\pi}{4}\right)$$

We introduce $i\pi/4$ so as to facilitate the transition through the left turning point x_1. Making use of the correspondence expressed by (31.20a), we see that the wave function below the barrier has the required form of a damped exponential:

$$|p|^{1/2}\psi = \exp\left(-\frac{1}{h}\int_{x_1}^{x} |p|\,dx\right)$$

To pass through the second rotation point x_2, we represent the function below the barrier as follows:

$$|p|^{1/2}\psi = \exp\left(-\frac{1}{h}\int_{x_1}^{x_2} |p|\,dx\right) \times \exp\left(\frac{1}{h}\int_{x}^{x_2} |p|\,dx\right) \quad (31.21)$$

Here, the first factor is a constant quantity, while the second increases exponentially into the barrier. A solution of this form is matched with a wave travelling from the barrier to the right and having a coefficient equal to modulus unity.

It follows from all that has been said that the amplitude of a wave receding beyond the barrier decreases in the ratio $\exp\left[-(1/h)\int_{x_1}^{x_2} |p|\,dx\right]$, while the probability of penetrating the barrier is equal to the square of the decrease in amplitude:

$$D = \exp\left(-\frac{2}{h}\int_{x_1}^{x_2} |p|\,dx\right) \quad (31.22)$$

This formula[9] can be used only as long as the exponent in it is large in comparison with unity. Otherwise, to develop the corresponding relationship it is necessary to use exact wave functions.

---

[9] Rigorous proof that the factor multiplying the exponential is unity is extremely involved. We have limited ourselves to a simplified derivation of (31.22).

The existence of tunneling shows that the concept of path is sometimes totally inapplicable to quantum motion. A path continued below the barrier would lead to imaginary values of the velocity.

The uncertainty relations (22.4a) indicate only the lower limit of possible inaccuracies in stating the coordinates and momenta. When a particle is located below the barrier, the inaccuracies increase greatly. Wherever it is located below the barrier, its velocity is an imaginary quantity, that is, it is completely indeterminate.

The same can be stated differently. In a change in the energy of a particle located in the domain below the barrier, the inaccuracy in the energy value is so great that we can no longer assert that the particle possesses energy there.

Alpha-Decay. Penetration of a potential barrier enables us to explain one of the most important facts of nuclear physics, alpha-decay. The nuclear masses of heavy elements with atomic numbers greater than that of lead satisfy an inequality of the form (14.9):

$$m\,(A,\,Z) > m\,(A - 4,\,Z - 2) + m\,(4,\,2)$$

Here A is the atomic weight and Z is the nuclear number. Thus, $m\,(4,\,2)$ is the mass of a helium nucleus with atomic weight 4 and atomic number 2. Such a nucleus emitted in alpha-decay is called an *alpha-particle.*

All that can be seen from the inequality is that the spontaneous decay of a nucleus of mass $m\,(A,\,Z)$ is possible, though no indication is obtained about the time law of disintegration. The nuclei of certain elements have mean decay times of 10^{10} years while others have decay times of about 10^{-5} s, which is a difference of 23 orders of magnitude. It will be noted that the energy of the alpha-particles emitted differs here by a factor of only two. From experiment it turns out that the logarithm of the mean decay time of a nucleus is inversely proportional to the alpha-particle velocity. It is this logarithmic law that corresponds to the difference of 23 magnitudes. It is accounted for by the difference of barrier factors which depend exponentially upon the energy.

The problem is to develop a suitable, and as far as possible simple, model of a nucleus to which the observed law of decay, based on quantum mechanical laws, would correspond. Such a model should not, furthermore, contradict other facts regarding the nucleus.

Although at present there is no quantitative theory of nuclear interactions, there exists a quite satisfactory, sufficiently universal model of a nucleus, which makes it possible to describe and reveal the interconnections of all observed phenomena in nuclear physics, up to particle energies of hundreds of MeV.

In the case of alpha-decay, the most general aspects of this model are sufficient. At large distances from the nucleus, exceeding 10^{-12} cm,

no specific nuclear interactions manifest themselves. At such distances an alpha-particle is subject only to the action of the electrostatic Coulomb force, to which corresponds the potential energy

$$U = \frac{2\,(Z-2)\,e^2}{r} \tag{31.23}$$

At small distances, attractive forces must act, of course, since otherwise the nucleus (A, Z) could not exist at all. We do not know the force law, that is, the shape of the potential-energy curve close to the nucleus, but we can assert that it should correspond to a very

$$U = \frac{2(Z-2)e^2}{r}$$
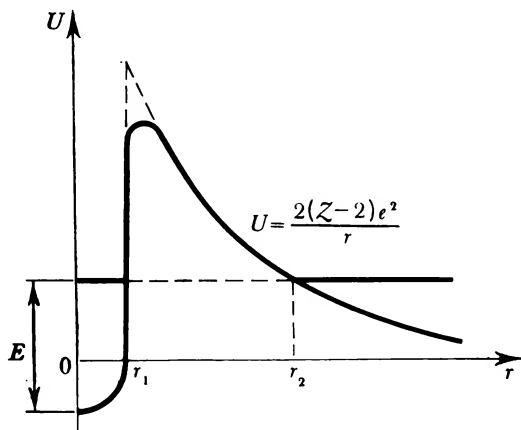
Figure 36

steep dependence of the forces on distance, as in all nuclear interactions. It was pointed out in Section 28 that simplified models known as potential wells are quite sufficient for the description of such potential curves.

Figure 36 presents such an assumed potential well in which an alpha-particle is located in the nucleus. Farther from the nucleus the curve corresponding to the well transforms in some way into a Coulomb repulsion curve; the finer details of the transformation law, as will be apparent from subsequent computations, are immaterial. The energy level E is plotted above zero, since otherwise the alpha-particle would simply be incapable of flying out of the well. The simple curve in Figure 36 is quite adequate to explain the law of alpha-decay.

To determine the probability of alpha-decay it is sufficient to calculate the barrier factor D from (31.22). Since nuclear forces are short-range, the transition region between electrostatic and nuclear forces is small, and the Coulomb law can be considered to hold up to the point $r = r_1$, that is, up to the vertical boundary of the well.

Point r_1 is the effective radius of the nucleus, determined from alpha-decay. (Other nuclear data lead to somewhat different values of its radius, but the differences have been found to be fully within acceptable limits.)

Thus, we compute the barrier factors from the formula

$$D = \exp\left\{ -\frac{2}{h} \int_{r_1}^{r_2} \left[2m\left(\frac{2(Z-2)e^2}{r} - E \right) \right]^{1/2} dr \right\} \equiv e^{-P}$$

(31.24)

The integral in the exponent can easily be calculated by the substitution

$$\frac{rE}{2(Z-2)e^2} = \cos^2 x$$

(31.25)

Then after elementary treatment it is reduced to

$$P = \frac{2}{h}(2m)^{1/2} \int_{r_1}^{r_2} \left(\frac{2(Z-2)e^2}{r} - E \right)^{1/2} dr$$

$$= \frac{2}{h}(2m)^{1/2} \frac{2(Z-2)e^2}{E^{1/2}} \left[\arccos\left(\frac{Er_1}{2(Z+2)e^2} \right)^{1/2} \right.$$

$$\left. - \left(\frac{Er_1}{2(Z-2)e^2} \right)^{1/2} \left(1 - \frac{Er_1}{2(Z-2)e^2} \right)^{1/2} \right] \quad (31.26)$$

The quantity $Er_1\,[2\,(Z-2)\,e^2]^{-1}$ is the ratio of the energy of the alpha-particle to the effective barrier height at point r_1, taken according to Eq. (31.23). Let us evaluate this ratio. For heavy nuclei, $2(Z-2) \approx 180$, $r_1 \approx 9 \times 10^{-13}$ cm, and we take E equal to 6 MeV (10^{-5} erg), $e^2 \approx 23 \times 10^{-20}$ esu^2. From this

$$\frac{Er_1}{2(Z-2)e^2} \approx \frac{1}{5}$$

This quantity can be considered small. Then in the right-hand side of (31.26) we obtain, approximately,

$$P \approx \frac{2}{h}(2m)^{1/2} \frac{2(Z-2)e^2}{E^{1/2}} \left[\frac{\pi}{2} - 2\left(\frac{Er_1}{2(Z-2)e^2} \right)^{1/2} \right]$$

$$\approx \frac{2\pi e^2}{hv} 2(Z-2) - \frac{8}{h}[mr_1 e^2 (Z-2)]^{1/2}$$

(31.27)

The validity of this expansion is easily checked by direct substitution.

Knowing the barrier factor, it is simple to find the time law of alpha-decay. For that we make use of Eq. (23.18). The surface integral in the right-hand side of this equation should be referred to an infinitely remote surface, insofar as outside the nucleus all

alpha-particles are receding from it. The total flux across such a sur-
face yields the probability of decay in unit time. The space integral
in the left-hand side of (23.18) need be extended only over the volume
of the nucleus, since the wave function of an alpha-particle falls off
exponentially below the barrier. According to the principal result
of Section 23, the probability of finding an alpha-particle in the
nucleus is

$$N = \int |\psi|^2 \, dV$$

We can assume that at some initial time N was equal to unity.
Then the law according to which it decreases with time can be de-
termined from Eq. (23.18). The amplitude of the wave function
decreases $D^{1/2}$ times in penetrating the barrier. If the amplitude
is assumed to be unity in the nucleus, as we did in calculating the
barrier factor, then the wave function at infinity should be

$$\psi = \frac{BD^{1/2}}{(4\pi)^{1/2}} \frac{e^{ipr/h}}{r} \tag{31.28}$$

We have made use of the fact that, to pass from one-dimensional
to three-dimensional motion, the wave function must be divided
by r (see Sec. 29). The factor $(4\pi)^{-1/2}$ appears in normalizing the
wave function to unity over the solid angle, the coefficient B is
associated with normalization to N over the volume of the nucleus.
Substituting the expression (31.28) into the right-hand side of
(23.18), which expresses the law of conservation of the number of
particles in nuclear decay, and replacing the normalization coef-
ficient B by the probability of finding the alpha-particle in the
nucleus, we can obtain the following law of alpha-decay:

$$\frac{dN}{dt} = -\frac{\Gamma}{h} N, \quad \Gamma = \frac{4E_i h}{mr_1 v} \frac{1}{[2(Z-2)e^2/(Er_1)-1]^{1/2}} e^{-P} \tag{31.29}$$

where E_i is the distance from the E energy level to the "bottom"
of the potential well in Figure 36; E_i is evaluated extremely roughly.
But the main significance of Eq. (31.29) lies not in the pre-exponen-
tial factor, but in the exponent of the exponential, which gives the
fundamental dependence of the decay probability upon the energy
of the alpha-particle. Taking this into account, and bearing in mind
the arbitrariness in the definition of E_i, we did not reproduce the
detailed development of the coefficient multiplying the barrier
factor in (31.29). This coefficient is entirely different if it is assumed
that an alpha-particle does not exist in the nucleus as an entity and
forms only at the moment of emission, which is probably closer to
the truth. But with this assumption, too, the barrier factor remains
the same.

Integrating (31.29), we obtain an exponential law for decreasing decay activity:

$$N = e^{-\Gamma t/h} \tag{31.30}$$

Here, we put $N(0) = 1$. The quantity has the dimensions of energy for the sake of convenience in comparing it with other quantities having such dimensions. Every nucleus has the same decay probability per unit time regardless of how long it has existed without disintegrating. This probability is Γ/h, and it is independent of time.

Equations (31.27) and (31.29) confirm the law of inverse proportionality between the logarithm of the probability of alpha-decay and the experimentally determined velocity v of an emitted α-particle. A simple computation shows that, for a twofold change in the energy of an α-particle, from 4 to 8 MeV, the disintegration time changes by 22 orders of magnitude. The laws of classical mechanics are quite incapable of explaining such a strong energy dependence of the decay time. But the cause of this dependence lies precisely in the quasi-classical motion of an α-particle: by classical laws it cannot leave the nucleus at all, while thanks to the finite value of the action quantum h there appears a very low probability, which decreases sharply together with the energy of the alpha-particle.

The Width of a Level. Using the expression (31.30), let us write in explicit form the time dependence of the wave function of a nucleus that has not emitted an alpha-particle. This dependence has the form

$$\psi \propto e^{-\Gamma t/(2h)} e^{-iEt/h} \tag{31.31}$$

The first factor accounts for the exponential fall-off of amplitude according to $e^{-\Gamma t/(2h)}$ law (since the probability, or the square of the amplitude, diminishes according to $e^{-\Gamma t/h}$); the second factor is the usual wave-function time factor. Expression (31.31) is very similar to the well-known formula for damped oscillations, with the difference that in the given case it is the probability amplitude of the initial (not yet decayed) state of the nucleus that is damped.

For this state there exists a finite probability flux for the emission of a particle from the nucleus, which is proportional to the probability of an alpha-particle being in the nucleus. It is this that leads to the exponential fall-off of the probability of the state prior to decay.

All nuclei before decay are described by exactly the same wave function (31.31), if at the instant $t = 0$ they were in the initial state. Therefore, they all have a perfectly identical probability of decaying in unit time, and it is impossible to predict which one of them will decay earlier and which later. In exactly the same way, in the diffraction experiment it is impossible to say which part of the

photographic plate will be hit by a given electron. The decay law is purely statistical, in the same way as the law for diffraction patterns.

Alpha-decay cannot be treated as an end result of some temporal process inside a nucleus: a nucleus is always, to exactly the same extent, ready for a decay process. This is indicated by the fixed form of the decay law.

For this reason time can, in principle, be equally measured by periodic or radioactive processes. Actually, the law of both processes is exponential: in one case it is with a real exponent, in the other, with an imaginary one.

The wave function (31.31) can be ascribed to the complex energy eigenvalue $E_c = E - i\Gamma/2$. This eigenvalue does not contradict the Hermiticity of the Hamiltonian. We noted in Section 26 that any differential operator is defined only after the boundary conditions imposed on the wave function are stated. Up till now we chose these conditions in real form: for example, for the case of finite motion the wave function was supposed to be zero in infinity. For the case of alpha-decay, we have a different boundary condition: at infinity the wave function is of the complex form (31.28) corresponding to a diverging spherical wave. The eigenvalue is complex because the wave function is complex.

Suppose now that the wave function (31.31) is expanded in a Fourier integral over real-energy functions whose time dependence is determined by the exponentials $e^{-iE't/h}$. What is the order of magnitude of the energy interval in which the amplitudes of the real-energy functions are other than zero?

We write the Fourier-integral expansion

$$e^{-iE_c t/h} \equiv e^{-\Gamma t/(2h) - iEt/h} = \int_{-\infty}^{\infty} a(E') e^{-iE't/h} dE' \qquad (31.32)$$

Then the amplitude $a(E')$ is

$$a(E') = \frac{1}{2\pi h} \int_0^{\infty} dt\, e^{-\Gamma t/h - i(E-E')t/h}$$

$$= \frac{1}{2\pi [\Gamma/2 + i(E-E')]} \qquad (31.33)$$

and the square of the amplitude is

$$|a(E')|^2 = \frac{1}{4\pi^2 [(E-E')^2 + \Gamma^2/4]} \qquad (31.34)$$

We see from this that $|a(E')|^2$ decreases by half when E' is at a distance $\pm\Gamma/2$ from E, so that the whole "half-width" of the energy interval is equal to Γ (Figure 37).

It can be seen from definition (31.32) that the dimensions of the amplitude $a(E')$ are (energy)$^{-1}$. Therefore, if we integrate the square of the amplitude over energy we obtain a quantity whose dimensions are also (energy)$^{-1}$. But this is the total area of the curve (31.34) characterizing the energy interval in which the amplitudes of the
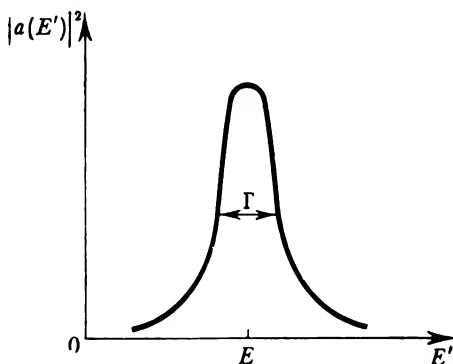
Figure 37

expansion are other than zero. Denoting this interval ΔE, we obtain, by definition

$$\frac{1}{\Delta E} \int_{-\infty}^{\infty} |a(E')|^2 \, dE' = \frac{1}{4\pi^2} \int_{-\infty}^{\infty} \frac{dE'}{(E-E')^2+\Gamma^2/4} = \frac{1}{2\pi\Gamma} \quad (31.35)$$

On the other hand, the mean decay time can be defined as

$$\Delta t = \int_{0}^{\infty} e^{-\Gamma t/h} t \, dt = \frac{h}{\Gamma} \quad (31.36)$$

Comparing (31.35) and (31.36) yields

$$\Delta E \, \Delta t = 2\pi h \quad (31.37)$$

The obtained relation is similar in form to the uncertainty relation (22.4a). Note that the factor 2π in the right-hand side of (31.37) is linked with the definition of ΔE through Eq. (31.35). If ΔE were determined as in (19.6b), according to the curve $|a(E')|$, in the right-hand side of (31.37) we would have unity.

The relation (31.37) should be formulated as follows: the energy of a state existing for a limited time Δt is determinate to the accuracy of a quantity of the order $2\pi h/\Delta t$. Only the energy of a state of infinite duration is defined exactly.

The meaning of the uncertainty relation for position and momentum is not analogous to the meaning of (31.37). The evaluation (22.4a) expresses the fact that position and momentum do not exist in the same state; (31.37) means that if the state of a system is of finite duration, Δt, then its energy at every instant belonging to Δt is not defined precisely and lies within some interval of values, ΔE, of the order $2\pi h/\Delta t$.

The quantity $\Gamma \sim \Delta E$ is the *width of the energy level* of the system. The concept of level width can be applied to any state of finite duration, not only to the states of systems capable of alpha-decay. For example, the energy level of an atom in an excited state has a definite nonzero width, since an excited atom is capable of spontaneous emission of a quantum.

Explanation of the Level Width. We shall now show how the level width of a nucleus capable of alpha-decay can be found by considering the wave function variation below a potential barrier.

It was shown in Section 28 that infinite motion has a continuous energy spectrum. The motion of a particle penetrating a potential barrier is infinite because it is capable of going to infinity. It follows, strictly speaking, that a nucleus capable of alpha-decay should have a continuous energy spectrum. Actually, we made use of this in the expansion (31.32).

Let us now see how the quasi-discrete levels E_c are found. From (31.32) we see that even for a nucleus with a very short alpha-decay time ($t \sim 10^{-5}$ s), $\Gamma \sim 10^{-22}$ erg or 0.6×10^{-10} eV. How is it possible to combine a continuous spectrum with such a narrow energy interval?

The general solution to the wave equation between points r and r_1 is of the following form:

$$\psi = \frac{C_1}{|p|^{1/2}} \exp \left(-\frac{1}{h} \int_{r_1}^{r} |p| \, dx \right) + \frac{C_2}{|p|^{1/2}} \exp \left(\frac{1}{h} \int_{r_1}^{r} |p| \, dx \right)$$

(31.38)

The first term exponentially decreases with r, while the second exponentially increases. It follows that if the barrier were extended to infinity rightwards, a solution would exist only for $C_2 = 0$. The ratio C_2/C_1, determined from the boundary conditions at $r = r_1$, is a function of energy. It is the roots of equation $C_2 (E) = 0$ that give the possible energy eigenvalues for finite motion. The energy of a particle in a well of finite depth is obtained in just this way.

Motion of a particle in a well, which was considered in Sec. 28, differs from motion of a particle beyond the barrier because the barrier is of finite width. Therefore, in this case the second solution,

proportional to C_2, need not be strictly equal to zero and may be just small compared with the first solution in any small interval of values, ΔE, close to a root of equation $C_2(E) = 0$. This region of values, ΔE, is what corresponds to the assumption that the modulus of the wave function outside the nucleus is small compared with the wave function inside the nucleus.

In other words, if the energy of the nucleus is contained in a given region of values ΔE, we can say that the alpha-particle is in a sense bound in the nucleus, that is, with overwhelming probability found within it over a certain finite time interval Δt.

The higher or wider the potential barrier, the less the barrier factor D and the less the decay probability Γ/h proportional to it. But then ΔE is also correspondingly reduced, that is, the continuous-spectrum state becomes closer to the discrete-spectrum state with an exact energy value E. This is what explains the meaning of the uncertainty ΔE: it indicates how close the state is to a bound one, with an infinite lifetime.

We can also say that as the barrier width increases, that is, as the barrier factor decreases, the imaginary part of the energy, equal to $-\Gamma/2$, tends to zero, while E_c tends to the bound-state energy.

The uncertainty ΔE in no way limits the applicability of the energy conservation law: the total energy of a nucleus and an alpha-particle is constant. But a state with a strictly defined energy cannot refer to an alpha-particle in a nucleus, since if its coordinate is given in an interval close to r_1, the energy can no longer have an exact value. As was pointed out before, this case is quite unlike free motion, therefore the uncertainty in energy should be calculated with the help of the decay probability, not simply from the relation (22.4a). If the energy is stated precisely, it refers to both an undecayed nucleus and a decayed nucleus. Superimposed, these states of the nucleus yield a general state with an exact energy value.

Any state capable of spontaneous transition to another state with the same energy possesses a certain energy width. A precisely defined energy always corresponds to a superposition of states capable of transition from one to the other.

We can divide the total level width into partial widths related to the probabilities for various transitions. Thus, strongly excited nuclear states are capable of emitting neutrons of various energies and of radiating gamma quanta. Each possibility contributes its exponential in the term characterizing attenuation of the wave function (31.31). The total attenuation is determined by the product of such exponentials. It follows that the total level width is equal to the sum of its widths in relation to all decay possibilities.

The Bohr Quantum Conditions. The quasi-classical approximation makes it possible to determine the energy levels of a particle

moving in a potential well. To avoid repetition of similar diagrams, we shall assume that in Figure 35 the bulge of the potential energy curve faces downwards. Then, between points x_1 and x_2 the total energy is greater than the potential energy, so that this is a classically possible domain of motion. According to the laws of classical mechanics, the particle performs periodic motion similar to a pendulum (cf. Fig. 8). The total period of the motion is

$$\tau = 2 \int_{x_1}^{x_2} \frac{dx}{v} = 2 (2m)^{1/2} \int_{x_1}^{x_2} \frac{dx}{[E - U(x)]^{1/2}} \qquad (31.39)$$

Let us now determine the restrictions imposed by quantum mechanics on the possible energy eigenvalues of the particle. Since to the right of point $x = x_2$ the potential energy is greater than the total energy, the solution of the wave equation must be represented in the form of an attenuating exponential:

$$\psi = \frac{1}{2 |p|^{1/2}} \exp \left(-\frac{1}{h} \int_{x_2}^{x} |p| dx \right)$$

Using the conditions for matching, (31.20), we should pass from this solution to the solution inside the well. We represent it as

$$\psi = \frac{1}{p^{1/2}} \exp \left(\int_{x}^{x_2} p \, dx + \frac{\pi}{4} \right) \qquad (31.40a)$$

since for x close to x_2 it transforms into (31.20a).

To the left of point $x = x_1$ the solution should again have the form of attenuating exponential:

$$\psi = \frac{C}{2 |p|^{1/2}} \exp \left(-\frac{1}{h} \int_{x_1}^{x} |p| dx \right)$$

If we apply the conditions for matching to it, the function in the region of the well will be expressed as follows:

$$\psi = \frac{C}{p^{1/2}} \sin \left(\frac{1}{h} \int_{x_1}^{x} p \, dx + \frac{\pi}{4} \right) \qquad (31.40b)$$

The functions (31.40a) and (31.40b) must, apparently, coincide for any value of x within the well, since they refer to the same state:

$$\sin \left(\frac{1}{h} \int_{x}^{x_2} p \, dx + \frac{\pi}{4} \right) = C \sin \left(\frac{1}{h} \int_{x_1}^{x} p \, dx + \frac{\pi}{4} \right) \qquad (31.41)$$

For Eq. (31.41) to be valid, one of two conditions must hold:
(i) $C = 1$, the sum of the phases under the sines is an even number multiplied by π.
(ii) $C = -1$, the sum of the phases is an odd number multiplied by π.
Indeed, in case (i), denoting the sine phases α and $2k\pi - \alpha$, we obtain

$$\sin \alpha - \sin (2k\pi - \alpha) = -2 \sin k\pi \cos (\alpha - k\pi) = 0$$

In case (ii) we have, similarly

$$\sin \alpha + \sin [(2k+1) \pi - \alpha]$$
$$= 2 \cos \left(k + \frac{1}{2}\right) \pi \sin \left[\alpha - \left(k + \frac{1}{2}\right) \pi\right] = 0$$

Combining both cases in one formula, we see that functions (31.40a) and (31.40b) coincide if the sum of the phases under the sine is equal to an integral of π which, evidently, is no less than π, because the sum of the phases is positive:

$$\frac{1}{h} \int\limits_{x}^{x_2} p \, dx + \frac{\pi}{4} + \frac{1}{h} \int\limits_{x_1}^{x} p \, dx + \frac{\pi}{4} = \pi (n + 1)$$

where $n = 0, 1, 2$, etc. Now, extending the integral over the whole period of motion in the classical sense, we arrive at the condition

$$2 \int\limits_{x_1}^{x_2} [2m (E_n - U)]^{1/2} \, dx \equiv 2 \int\limits_{x_1}^{x_2} p_n \, dx = 2\pi \left(n + \frac{1}{2}\right) h \quad (31.42)$$

A similar condition was postulated in Bohr's old theory, but then simply n was written. Note that in contemporary quantum mechanics the terms $\pi/4$ in the cosine phases appear only when the potential energy curves have finite slopes at the turning points. If that is not the case, further investigation of the question of the correct choice of phases is required. In general, the quasi-classical approximation is applicable only when the phases involved in the wave function are large in comparison with 2π. Formula (31.42) is therefore valid only for large n's. There are, however, exceptional cases when it is applicable for all n's.

As distinct from the old quantum theory, quantum mechanics obtains the approximate formula (31.42) without any additional, extraneous, assumptions, whereas formerly such a formula was applied to classical motion. But in classical mechanics it is always assumed that mechanical quantities may vary continuously, so that the quantum postulate appears sharply to contradict all its principles.

As can be seen from (31.42), the phase of the wave function on the segment (x_1, x_2) varies by a number greater than πn and smaller than $\pi(n + 1)$. It follows that the sine changes its sign n times, that is, the function has n zeros, which corresponds to the nth energy eigenvalue. The eigenvalues E_n increase together with the number n, because only on that condition can the integral $\int_{x_1}^{x_2} (E_n - U)^{1/2}\, dx$, that is, the area of the curve, $\int_{x_1}^{x_2} p\, dx$, lying below the line $E_n =$ $=$ constant, increase. Thus, the general rule is confirmed: the greater the energy eigenvalue the greater the number of zeros in the corresponding wave function.

Let us find the interval ΔE between two neighbouring energy eigenvalues for large n's. We have

$$2 \int_{x_1}^{x_2} \frac{\Delta E \,(2m)^{1/2}}{[E_n - U(x)]^{1/2}}\, dx = 2\pi h \tag{31.43}$$

But the integral here is expressed in terms of the classical motion period τ in such a way that

$$\Delta E = \frac{2\pi h}{\tau} = h\omega \tag{31.44}$$

where ω is the frequency of classical motion. Thus, in the first approximation successive energy levels are equidistant, if we can neglect the dependence of the oscillation frequency on the energy.

The methods of finding the eigenvalues of the Hamiltonian described here can also be applied to other operators, for example, to the angular momentum square. For that the equation for the wave function of such an operator must be reducible to the form

$$f'' + p^2(x, \lambda)\, f = 0$$

where λ is the eigenvalue of the investigated operator.

The Quasi-Classical Limit of Matrix Elements. Let us now see what the matrix elements of operators transform into in the quasi-classical approximation. For a limiting process from the wave laws of motion of quantum mechanics to classical motion along paths, we must form wave packets from the wave functions:

$$\psi = \int dE\, C(E)\, \psi(E, x), \qquad \psi^* = \int dE'\, C^*(E')\, \psi^*(E', x) \tag{31.45}$$

Since actually the classical path is considerably smeared in comparison with quantities on the microscopic scale, the energy interval, ΔE, over which the integration is performed in (31.45) is very small.

Let there be a certain operator $\hat{\lambda}(x)$ from which we must develop the average value over the state describing the wave packet (31.45). In accordance with the general formula (25.19), we write

$$\langle \lambda(t) \rangle = \int \bar{\psi}^* \hat{\lambda} \bar{\psi} \, dx$$

$$= \int dE \, C(E) \int dE' \, C^*(E') \, e^{i(E'-E)t/h} \lambda_{E'E} \qquad (31.46)$$

We take advantage of the fact that the whole energy interval in the quasi-classical limit is very small and replace $E' - E$ by the quantity ΔE to get

$$\langle \lambda(t) \rangle = \int \int dE \, d\Delta E \, C(E) \, C^*(E + \Delta E) \, e^{i\Delta Et/h} \lambda_{E+\Delta E \, E}$$

In $C^*(E + \Delta E)$ we can simply replace ΔE by 0. Then $C(E) \, C^*(E + \Delta E)$ transforms into $|C(E)|^2$, that is, the probability of the state per unit energy. In the matrix element $\lambda_{E'E}$ such a substitution would mean passing to a diagonal element, which in general does not depend upon time, whereas we are interested in $\langle \lambda(t) \rangle$ as a function of t.

The matrix element $\lambda_{E+\Delta E \, E}$ depends on the energy E smoothly. It is possible to substitute into it some mean value of the energy E_0 in the interval ΔE. Actually, $\langle \lambda \rangle = \langle \lambda(E_0, t) \rangle$ also depends upon E_0. Then the required mean value is expressed as follows:

$$\langle \lambda(E_0, t) \rangle = \int dE \, |C(E)|^2 \int d\Delta E \, e^{i\Delta Et/h} \lambda_{E_0+\Delta E \, E_0} \qquad (31.47)$$

But the integral of all the probabilities, $\int dE \, |C(E)|^2$, is unity, and we come to the following:

$$\langle \lambda(E_0, t) \rangle = \int d\Delta E \, e^{i\Delta Et/h} \lambda_{E_0+\Delta E \, E_0} \qquad (31.48)$$

If we write the expansion of the classical quantity $\lambda(E_0, t)$ in a Fourier integral, a simple comparison of (31.48) and the expansion

$$\lambda(E_0, t) = \int d\omega \, e^{i\omega t} l(E_0, \omega)$$

shows that

$$\lambda_{E_0+\Delta E \, E_0} = h^{-1} l(E_0, \Delta E/h) \qquad (31.49)$$

Thus, the matrix element transforms into a coefficient of the Fourier expansion corresponding to the frequency with which the matrix itself varies with time. In the case of a discrete spectrum, we should employ not the integral but a Fourier series, the principal interval of the expansion being the period of classical motion, τ, determined from (31.39).

Let us now apply the obtained result to show that the commutator between the Hamiltonian and the operator $\hat{\lambda}$ in the classical limit passes into the Poisson bracket for the corresponding functions $\mathcal{H}(x, p)$ and $\hat{\lambda}(x, p)$ (see Sec. 27).

Let the operators $\hat{\mathcal{H}}$ and $\hat{\lambda}$ be in coordinate representation. Then their commutator looks like this:

$$(\hat{\mathcal{H}}\hat{\lambda} - \hat{\lambda}\hat{\mathcal{H}})_{xx'}$$
$$= \int dx'' [\mathcal{H}(x, x'')\lambda(x'', x') - \lambda(x, x'')\mathcal{H}(x'', x')]$$

Every function of two variables can be identically represented as a function of the sum of, and difference between, those variables:

$$\mathcal{H}(x, x'') \equiv \mathcal{H}\left(\frac{x + x''}{2}, x - x''\right)$$
$$\lambda(x'', x') \equiv \lambda\left(\frac{x'' + x'}{2}, x'' - x'\right)$$
$$\lambda(x, x'') \equiv \lambda\left(\frac{x + x''}{2}, x - x''\right)$$
$$\mathcal{H}(x'', x') \equiv \mathcal{H}\left(\frac{x'' + x'}{2}, x'' - x'\right)$$

We expand the functions of the difference in the right-hand sides of these identities in Fourier integrals in terms of the normalized momentum eigenfunctions, $\psi = (2\pi h)^{-1/2} e^{ipx/h}$, to get

$$\mathcal{H}(x, x'') = \frac{1}{2\pi h} \int e^{ip(x-x'')/h}\mathcal{H}\left(\frac{x + x''}{2}, p\right) dp$$
$$\lambda(x'', x') = \frac{1}{2\pi h} \int e^{iT_1(x''-x')/h}\lambda\left(\frac{x'' + x'}{2}, p_1\right) dp_1$$

In the classical limit, as was pointed out before, we must pass to wave packets and assume the differences between the coordinates to be small quantities. We therefore expand $\mathcal{H}\left(\frac{x + x''}{2}, p\right)$, $\lambda\left(\frac{x'' + x'}{2}, p_1\right)$ and the other Fourier amplitudes in power series in $x - x''$, $x' - x'$, ..., and restrict ourselves to linear terms:

$$\mathcal{H}(x, x'') = \frac{1}{2\pi h} \int e^{ip\Delta x/h}\left(\mathcal{H}(x, p) + \frac{1}{2}\Delta x \frac{\partial}{\partial x}\mathcal{H}(x, p)\right) dp$$
$$\lambda(x'', x') = \frac{1}{2\pi h} \int e^{ip_1\Delta x''/h}\left(\lambda(x', p) - \frac{1}{2}\Delta x'' \frac{\partial}{\partial x'}\lambda(x', p_1)\right) dp_1$$

where $\Delta x = x - x''$, and $\Delta x'' = x'' - x'$. Next we transform the linear terms by parts. For this we replace $\Delta x e^{ip\Delta x/h}$ by $(h/i) \times \times (\partial/\partial p) e^{ip\Delta x/h}$ and take advantage of the fact that at the limits

(for $p = \pm\infty$) the expansion coefficients must vanish. Hence

$$\mathcal{H}(x, x'') = \frac{1}{2\pi h} \int e^{ip\Delta x/h} \left(\mathcal{H}(x, p) - \frac{h}{2i} \frac{\partial^2 \mathcal{H}}{\partial x \, \partial \mu} \right) dp$$

$$\lambda(x'', x') = \frac{1}{2\pi h} \int e^{ip_1\Delta x''/h} \left(\lambda(x', p_1) + \frac{h}{2i} \frac{\partial^2 \lambda}{\partial x' \, \partial p_1} \right) dp_1$$

(The signs of the second derivatives are opposite because the powers of the exponentials involve, respectively, $x - x''$ and $x'' - x'$.)

We substitute the expansions of the matrix elements into the commutator for the operators and make use of the fact that changing the order of integration with respect to x'', p, and p_1 yields a δ function:

$$\frac{1}{2\pi h} \int e^{ix''(p_1 - p)/h} \, dx'' = \delta(p_1 - p)$$

Substituting the integral over p_1 with the help of the formula

$$\int dp_1 \, \delta(p_1 - p) f(p_1) = f(p)$$

and retaining only the terms linear with respect to the action quantum, we arrive at the following expression for the required matrix element:

$$(\hat{\mathcal{H}}\hat{\lambda} - \hat{\lambda}\hat{\mathcal{H}})_{xx'}$$

$$= \frac{1}{2\pi h} \int e^{ip(x-x')/h} \, dp \left\{ \mathcal{H}(x, p) \lambda(x', p) - \mathcal{H}(x', p) \lambda(x, p) \right.$$

$$- \frac{h}{2i} \left(\frac{\partial^2 \mathcal{H}}{\partial x \, \partial \mu} \lambda(x', p) - \mathcal{H}(x, p) \frac{\partial^2 \lambda}{\partial x' \, \partial p} \right.$$

$$\left. \left. + \lambda(x, p) \frac{\partial^2 \mathcal{H}}{\partial x' \, \partial \mu} - \frac{\partial^2 \lambda}{\partial x \, \partial \mu} \mathcal{H}(x', P) \right) \right\}$$

It is significant that $\frac{\partial^2 \mathcal{H}}{\partial x \, \partial \mu}$ and $\frac{\partial^2 \mathcal{H}}{\partial x' \, \partial \mu}$ are now involved with the same signs, since the exponentials have, respectively, $x - x''$ and $x'' - x'$. We again transform by parts the terms proportional to h. As an example, let us consider one of them:

$$\int e^{ip(x-x')/h} \frac{\partial^2 \mathcal{H}}{\partial x \, \partial p} \lambda \, dp = - \int \frac{\partial \mathcal{H}}{\partial x} \frac{\partial}{\partial p} \left(\lambda e^{ip(x-x')/h} \right) dp$$

$$= - \int \left(\frac{i}{h} (x - x') \lambda \frac{\partial \mathcal{H}}{\partial x} + \frac{\partial \lambda}{\partial p} \frac{\partial \mathcal{H}}{\partial x} \right) e^{ip(x-x')/h} \, dp$$

Passing to the limit, we should assume that $x' \to x$ everywhere except, of course, the exponential $e^{ip(x-x')/h}$ with respect to which the expansion is performed. Then

$$\int e^{ip(x-x')/h} \frac{\partial^2 \mathcal{H}}{\partial x \, \partial p} \lambda(x', p) \, dp \to - \int e^{ip(x-x')/h} \frac{\partial \mathcal{H}}{\partial x} \frac{\partial \lambda}{\partial p} \, dp$$

Collecting all terms linear with respect to h, we obtain the commutator in the classical limit:

$$(\hat{\mathcal{H}}\hat{\lambda} - \hat{\lambda}\hat{\mathcal{H}})_{x' \to x} = \frac{1}{|2\pi h} \int dp\, e^{ip\,(x-x')/h}x$$

$$\times \frac{h}{i} \left(\frac{\partial \mathcal{H}}{\partial p} \frac{\partial \lambda}{\partial x} - \frac{\partial \mathcal{H}}{\partial x} \frac{\partial \lambda}{\partial p} \right) \quad (31.50)$$

We now expand (31.50) in a Fourier integral in the same way as we expanded $\dot{\mathcal{H}}$ and λ to get

$$(\hat{\mathcal{H}}\hat{\lambda} - \hat{\lambda}\hat{\mathcal{H}})_{x' \to x} = \frac{1}{2\pi h} \int dp\, e^{ip\,(x-x')/h} (\hat{\mathcal{H}}\hat{\lambda} - \hat{\lambda}\hat{\mathcal{H}})_{xp} \quad (31.51)$$

Here, we can directly assume the expression in parentheses to be a function of x and p. Comparing the coefficients of the Fourier expansion (31.51) with (31.50), we see that on the one hand we have the classical limit of a matrix element of the commutator $(\hat{\mathcal{H}}\hat{\lambda} - \hat{\lambda}\hat{\mathcal{H}})$, and on the other we have the Poisson bracket:

$$\lim_{[h \to 0} \frac{i}{h} (\hat{\mathcal{H}}\hat{\lambda} - \hat{\lambda}\hat{\mathcal{H}}) = \frac{\partial \mathcal{H}}{\partial p} \frac{\partial \lambda}{\partial x} - \frac{\partial \mathcal{H}}{\partial x} \frac{\partial \lambda}{\partial p} \quad (31.52)$$

as was pointed out in Section 27. Note, that the next term in the expansion of the commutator already involves h^2.

EXERCISES

1. Determine the energy levels of a linear harmonic oscillator from Eq. (31.42).

Solution.

$$\int_{-(2E/m\omega^2)^{1/2}}^{(2E/m\omega^2)^{1/2}} d\bar{x} \left[2m \left(E - \frac{m\omega^2 r^2}{2} \right) \right]^{1/2} = \frac{\pi E}{\omega} = \pi h \left(n + \frac{1}{2} \right)$$

From this, $E_n = h\omega\,(n + 1/2)$, which is in general valid for all n's, from zero on. Note that in a formula suitable only for large n's the term 1/2 is meaningless.

2. Determine the factor D for a potential barrier of the form: $U = 0$ for $x < 0$, $U = U_0 - \alpha x$ for $x \geqslant 0$, and $E < U_0$.

3. Compare the accidental energy degeneracy in the hydrogen-atom problem with the expression for energy in terms of the adiabatic invariants in Kepler's problem (Exercise 3, Section 10).

32

PERTURBATION THEORY

The Hamiltonian of a quantum mechanical system frequently involves a term multiplied by a small parameter. If the exact wave equation does not admit of an analytical solution, or if it is very complicated, it is useful to seek a solution in the form of an expansion in powers of the small parameter.

The problem can be posed in two ways. Sometimes the small term in the Hamiltonian, which is called the *perturbation*, but slightly affects the energy and the wave function of the unperturbed state, the perturbed and unperturbed states both being stationary.

In other cases the perturbation makes a stationary state non-stationary, so that a finite probability appears of the system passing into an entirely different state, for example, from one with discrete energy eigenvalues to one with a continuous energy spectrum.

These cases must be examined separately.

Time Independent Perturbations. Nondegenerate Motion. We shall first examine the simplest case, when the energy eigenvalues of the Hamiltonian corresponding to the unperturbed state of a system are not degenerate. Denoting the unperturbed Hamiltonian $\hat{\mathscr{H}}_0$, and the perturbation $\hat{V}$, the total Hamiltonian is

$$\hat{\mathscr{H}} = \hat{\mathscr{H}}_0 + \hat{V} \tag{32.1}$$

The equation for the eigenfunctions of the unperturbed Hamiltonian looks like this:

$$(\hat{\mathscr{H}}_0 - E_{n0})\psi_{n0} = 0 \tag{32.2}$$

We assume that $\hat{\mathscr{H}}_0$ has a discrete spectrum and that to each eigenvalue of E_{n0} there corresponds one eigenfunction ψ_{n0}, these functions being orthogonal and normalized to unity:

$$\int \psi_{n'0}^* \psi_{n0}\, dx = \delta_{n'n} \tag{32.3}$$

The perturbation involves a certain small parameter, but we shall not write it in explicit form, remembering that the order of each expression is determined by its degree with respect to $\hat{V}$.

We shall seek the correction to the eigenfunction and to the energy eigenvalue with number n. The corresponding exact equation for the energy eigenvalues is

$$(\hat{\mathscr{H}}_0 + \hat{V} - E_n)\psi_n = 0 \tag{32.4}$$

The solution is conveniently sought in the form of an expansion in eigenfunctions of the unperturbed Hamiltonian, ψ_{n0}, which are determined from (32.3):

$$\psi_n = \sum_{n'} C_{n'} \psi_{n'0} \tag{32.5}$$

Substituting this expansion into Eq. (32.4) and making use of the fact that $\psi_{n'0}$ are the eigenfunctions of $\hat{\mathscr{H}}_0$, we obtain

$$\sum_{n'} C_{n'} (E_{n'0} - E_n + \hat{V}) \psi_{n'0} = 0$$

This equation is exact. It is readily transformed into an equation for the coefficients $C_{n'}$, which corresponds to a transition to a representation in which the energy of the unperturbed system (32.2) is the independent variable. For this we premultiply the equation for ψ by ψ_{k0}^* and integrate over the whole volume (we denote the volume element dx so as to avoid confusion with the operator $\hat{V}$). Making use of the orthogonality of wave functions, we find

$$(E_{k0} - E_n) \, C_k = - \sum_{n'} V_{kn'} C_{n'} \tag{32.6}$$

Here V_{kn} denotes a matrix element of the perturbation.

The relation (32.6) represents an infinite set of linear equations with respect to the expansion coefficients $C_{n'}$. This set is solved by method of expansion in a power series in the small parameter involved in V. We represent the energy of the nth state in the form

$$E_n = E_{n0} + E_{n1} + E_{n2} + \ldots \tag{32.7}$$

and the coefficients of the expansion in the eigenfunctions of the unperturbed motion as

$$C_k = \delta_{kn} + C_{k1} + C_{k2} + \ldots \tag{32.8}$$

Now, in the left-hand side of (32.6) we put $k = n$; then in the first approximation there remains only one correction to the energy, E_{n1}. In the right-hand side we must retain the first-order term with respect to V, or V_{nn}. Hence, in the first approximation the equation for the correction to the energy of the unperturbed motion is

$$E_{n1} = V_{nn} \tag{32.9}$$

But the diagonal matrix element of the perturbation is

$$V_{nn} = \int \psi_{n0}^* \hat{V} \psi_{n0} \, dx \tag{32.10}$$

that is, in accordance with (25.19) the first-approximation correction to the energy is equal to the mean value of the energy for the unperturbed state.

If $k \neq n$, the difference $E_{k0} - E_n$ is of zero order with respect to the perturbation; it is also other than zero in the unperturbed system. Hence the whole left-hand side of Eq. (32.6) is of the first order, since it involves C_k $(k \neq n)$. To have the same order on the right as well, we must retain with $V_{kn'}$ only the coefficient with the number $n' = n$, in accordance with (32.8). From this we find C_{k1}:

$$C_{k1} = -\frac{V_{kn}}{E_{k0} - E_{n0}} \tag{32.11}$$

(we have replaced E_n by E_{n0} in the denominator, since we already have V_{kn} in the numerator).

With the help of (32.5) we find the eigenfunction (in the first approximation):

$$\psi_n = \psi_{n0} - \sum_k{}' \frac{V_{kn}\psi_{k0}}{E_{k0} - E_{n0}} \tag{32.12}$$

where the primed summation sign denotes that the term with $k = n$ has been discarded.

From Eqs. (32.11) and (32.12), we can see the requirement the perturbation must satisfy to be considered small:

$$|V_{kn}| \ll |E_{k0} - E_{n0}| \tag{32.13}$$

It should be small not with respect to the unperturbed energy eigenvalues, but with respect to their differences. The matrix element of the perturbation between the states k and n must be substantially smaller than the distance between the kth and nth energy levels.

Often the mean value of the perturbation with respect to the unperturbed state vanishes. Then the next approximation should be used. We again substitute $k = n$ into Eq. (32.6), but on the left retain the terms up to the second order. Taking into account that on the right we already have $V_{nn'}$, we must retain the coefficients $C_{n'}$ up to the first order:

$$E_{n1} + E_{n2} = V_{nn} + \sum_{n'}{}' V_{nn'}C_{n'1}$$

$$= -\sum_{n'}{}' \frac{V_{nn'}V_{n'n}}{E_{0n'} - E_{0n}} + V_{nn} \tag{32.14}$$

The quantities E_{n1} and V_{nn} cancel out in accordance with (32.9), leaving the expression for the correction to the energy of the nth state in the second approximation:

$$E_{n2} = -\sum_{n'}{}' \frac{|V_{nn'}|^2}{E_{0n'} - E_{0n}} \tag{32.15}$$

Here we made use of the fact that $V_{nn'} = V_{n'n}^*$, since $\hat{V}$ must be a Hermitian operator.

Note that if $n = 0$, that is, the ground state is being perturbed, the correction to the energy E_0 must necessarily be negative.

The Variation Property of Eigenvalues. This last result can be explained on the basis of very general considerations. We pointed out that finding the eigenvalues of an operator represents no more than a reduction to the principal axes of a second-order surface in Hilbert space. But the principal axes of a quadratic surface possess the variation property: in an infinitesimal rotation of the coordinate axes (close to the principal axes) the radius of a surface drawn from the origin of the coordinate system is stationary, that is (in the first approximation) it does not change. Consequently, if we speak of the zero, or ground, state, the corresponding principal axis is the smallest; any approach to this state leads only to a reduction in the energy.

The eigenvectors of $\hat{\mathcal{H}}_0$ did not yield the actual principal axes of the operator $\hat{\mathcal{H}} = \hat{\mathcal{H}}_0 + \hat{V}$. But when a correction to the wave functions, that is, the eigenvectors of $\hat{\mathcal{H}}$, was introduced with the help of (32.12), the eigenvalues came closer to the actual ones. In particular, the ground-state energy was brought closer to its actual value and, consequently, it decreased.

As for the correction to the first approximation, $E_{n1} = V_{nn}$, it can be of any sign. This correction is computed from the eigenfunctions of the unperturbed state, without rotating the axes in Hilbert space. In effect it should simply be included in the energy of the unperturbed state. It determines the correction to the radius of the quadratic surface of the exact Hamiltonian $\hat{\mathcal{H}} = \hat{\mathcal{H}}_0 + \hat{V}$ drawn in the direction of the nth principal axis of the unperturbed Hamiltonian. It is precisely this quantity, corrected in the first approximation, that is further refined by a rotation of the axis according to the formula (32.12), and it is then found that in a more exact refinement of the ground-state energy it can receive only a negative correction, E_{02}.

The variation property of eigenvalues can be very simply proved in the following way. We seek the wave function ψ, which gives the integral

$$\langle \hat{\mathcal{H}} \rangle = \int \psi^* \hat{\mathcal{H}} \psi \, dx \tag{32.16}$$

its extremal value, with the additional condition

$$1 = \int \psi^* \psi \, dx \tag{32.17}$$

that is, that it remains always normalized to unity.

To solve the extremum problem with the additional condition, we should employ the method of undetermined multipliers: vary both integrals (32.16) and (32.17) with respect to ψ^*, multiply the variation of the second integral by the as yet undeterminate quantity $-E$, add after this the variations of the integrals, and equate their sum to zero to get

$$\int \delta\psi^* (\hat{\mathscr{H}} - E)\,\psi\,dx = 0$$

For this equation to hold for any variation $\delta\psi^*$ we must require that

$$(\hat{\mathscr{H}} - E)\,\psi = 0 \tag{32.18}$$

But this is the equation for finding the eigenfunctions of $\hat{\mathscr{H}}$. Hence, any eigenvalue corresponds to the stationary value $\langle\mathscr{H}\rangle$, the ground state having the least possible value.

This property of the ground-state energy is used in the following way. For example, we have a neutral atom and an electron and have to determine whether they can join, forming a negative ion (electrons can be captured by atoms of hydrogen and several other elements).

It is extremely difficult to solve the problem of the motion of two electrons in a nuclear field. An analytical solution does not exist at all. This is where the variation method comes in. Certain eigenfunctions are chosen which satisfy the boundary conditions, that is, are equal to zero at infinity, as should be in the case of finite motion of a bound electron. These functions by no means satisfy the exact Schrödinger equation for the given problem and can only resemble it somewhat.

If substitution into (32.16) yields a negative value of $\langle\mathscr{H}\rangle$, we can be sure that the actual energy eigenvalue in the ground state lies even lower, that is, it is also negative. But negative energy eigenvalues indeed correspond to finite motion, or to the bound state of the electron. Usually the functions substituted into the expression for $\langle\mathscr{H}\rangle$ involve a certain number of parameters at the discretion of the person carrying out the computation. If the point of interest is the ground-state energy, these functions should never vanish at finite distances from the origin of the coordinate system, which somewhat restricts their choice. After the integrals (32.16) and (32.17) have been calculated, the parameters are so chosen as to have $\langle\mathscr{H}\rangle$ as small as possible while preserving the normalization of the function.

It should be noted that this method can be used only for determining, more or less accurately, the energy. The wave function may not resemble the actual function very much.

Degenerate Perturbation Theory. Suppose now that the nth energy eigenvalue in the unperturbed state is degenerate. There are several

functions $\psi_{n\lambda0}$ that satisfy the same equation

$$\hat{\mathscr{H}}_0\psi_{n\lambda0} = E_{n0}\psi_{n\lambda0} \tag{32.19}$$

Here λ may correspond to the eigenvalue of some operator that commutes with the Hamiltonian. Then it is necessary to choose correctly the eigenfunction of zero approximation, ψ_{n0}, which should be given the form

$$\psi_{n0} = \sum_\lambda a_\lambda\psi_{n\lambda0} \tag{32.20}$$

that is, it must be sought as a certain linear superposition of the initial functions. In this case the expansion (32.8) should be represented as follows:

$$C_{n'\lambda} = \delta_{nn'}a_{n\lambda} + C_{n'\lambda}^{(1)} + \ldots \tag{32.21}$$

The equation for finding E_{n1} changes accordingly:

$$E_{n1}a_{n\lambda} = \sum_{\lambda'} V_{n\lambda\,n\lambda'}a_{n\lambda'} \tag{32.22}$$

where $V_{n\lambda\,n\lambda'}$ is the matrix element between the states $n\lambda$ and $n\lambda'$:

$$V_{n\lambda\,n\lambda'} = \int \psi_{n\lambda}^*\hat{V}\psi_{n\lambda'}\,dx \tag{32.23}$$

The set of equations (32.22) has a solution if the determinant vanishes:

$$|V_{n\lambda\,n\lambda'} - \delta_{\lambda\lambda'}E_{n1}| = 0 \tag{32.24}$$

In the most general case, the number of roots of the determinant is equal to the degeneracy multiplicity of the unperturbed state ψ_{n0}. To each root there corresponds a definite eigenvalue $E_{n1}^{(\nu)}$ and a definite set of expansion coefficients $a_{n\lambda}$. Thus, the degenerate state splits into nondegenerate states; in the most general case their number is equal to the *degeneracy multiplicity*.

Let us explain this from the physical point of view. Suppose that a degenerate unperturbed state was of the required kind, that is, it was due to a certain symmetry in the statement of the problem. For example, in any central field, if the total angular momentum is not zero, degeneracy occurs with respect to the magnetic quantum number. If a perturbation violates central symmetry, then different energies of the system may correspond to different projections of the angular momentum on some axis. For example, in a magnetic field $\mathbf{H}$ a certain term $-\beta|\mathrm{H}|k$, where β is the Bohr magneton, is added to the energy of the atom. A splitting of states occurs, but it is possible to find the correction to the energy without solving Eq. (32.24). This is because in the present case the perturbation

caused by the magnetic field is

$$\hat{V} = -\beta \,|\, \mathbf{H} \,|\, \hat{M}_z = -\beta \,|\, \mathbf{H} \,|\, \frac{h}{i} \, \frac{\partial}{\partial \varphi} \qquad (32.25)$$

But the operator $(h/i)\,(\partial/\partial\varphi)$ is diagonal in the states with a given value of the magnetic quantum number k.

Perturbation as a Cause of Quantum Transitions. Consider the following problem. Let the Hamiltonian of an unperturbed system have states belonging to both discrete and continuous energy spectra, but such that they refer to the same energy values. We saw an example of such states in Section 29: an excited atom may have sufficient energy to disintegrate into a positive ion and a free electron, but the latter remains bound, since otherwise the law of conservation of parity would be violated. (We remember that if the Hamiltonian is an even function of the coordinates, total parity of the wave function is conserved in the system.)

Another example of such a state with a discrete spectrum capable of passing into a state with a continuous spectrum is an excited atom plus an electromagnetic field. As long as the atom does not interact with the field the excited state can exist indefinitely simply because the atom has nowhere to transfer its excitation energy. But if an external perturbation—the electromagnetic field—is made to act on the system (or if we have taken into account the known additional components of the Hamiltonian, which can be treated as a perturbation), the state with a discrete energy level is no longer strictly stationary.

Suppose that in the first example a certain perturbation, the operator of which is not an even function of the coordinates, acts on the atom. This violates the parity conservation law, and transition to a continuous spectrum is possible. In the second example, the operator of an interaction between an electron and the electromagnetic field is included in the Hamiltonian. This interaction makes possible the emission of a light quantum. But the energy of the quantum is equal to $h\omega$, and the frequency ω takes on a continuous set of values. Hence, the electromagnetic field energy has a continuous spectrum. Here, the initial state of the system is an excited atom and an electromagnetic field in the absence of quanta, while the final state is an atom in ground state and one quantum in the field.

The third example that should be cited is that of particle scattering on a force centre. Here, both particle states, the initial and the end, belong to a continuous spectrum, since the particle is free in them. If the scattering is elastic, the energy of the initial and final states of the particle is the same, and it is the direction of the momentum that changes. In this case the perturbation is due to the scattering centre, which transforms the particle from one state with a con-

tinuous spectrum to another but conserves its energy. In inelastic scattering a portion of the energy may be transmitted to the scatterer, but the important thing is that the particle's energy nevertheless continues to belong to a continuous spectrum.

Our task is to determine the probability of a perturbation causing the energy of a system to pass into the continuous spectrum.

Since we are dealing with a transition, we should proceed from the time dependent Schrödinger equation

$$-\frac{h}{i}\frac{\partial\psi}{\partial t}=(\hat{\mathscr{H}}_0+\hat{V})\,\psi \tag{32.26}$$

The corresponding equation for the unperturbed Hamiltonian has the form

$$-\frac{h}{i}\frac{\partial\psi_m^{(0)}}{\partial t}=\hat{\mathscr{H}}_0\psi_m^{(0)} \tag{32.27}$$

Assuming $\hat{V}$ to be a small perturbation, we represent the wave function as

$$\psi=\psi^{(0)}+\psi^{(1)} \tag{32.28}$$

and neglect the term $\hat{V}\psi^{(1)}$ as being of second order. Then for $\psi^{(1)}$ we obtain a nonhomogeneous equation

$$-\frac{h}{i}\frac{\partial\psi^{(1)}}{\partial t}-\hat{\mathscr{H}}_0\psi^{(1)}=\hat{V}\psi^{(1)} \tag{32.29}$$

We look for $\psi^{(1)}$ in the form of an expansion in the eigenfunctions of $\hat{\mathscr{H}}_0$:

$$\psi^{(1)}=\sum_m C_m(t)\,\psi_m^{(0)} \tag{32.30}$$

each of $\psi_m^{(0)}$ satisfying the homogeneous equation (32.27). Substituting (32.30) into the nonhomogeneous equation, and taking into account (32.27), we arrive at the equation

$$-\frac{h}{i}\sum_m\frac{\partial C_m}{\partial t}\,\psi_m^{(0)}=\hat{V}\psi_1^{(0)} \tag{32.31}$$

From this, with the help of the orthogonality property of $\psi_m^{(0)}$, we obtain the equations for the coefficients C_m. We multiply both sides of (32.31) by $\psi_n^{(0)*}$ and integrate over the volume. Then in the left-hand side there remains only the term $-(h/i)(\partial C_n/\partial t)$, and in the right-hand side the matrix element of the perturbing potential:

$$-\frac{h}{i}\frac{\partial C_n}{\partial t}=V_{n1}C_1 \tag{32.32}$$

We separate the time dependence explicitly:

$$V_{n1} = \int \psi_n^{(0)*} \hat{V} \psi_1^{(0)} \, dx = e^{i\,(E_n - E_1)\,t/h}$$
$$\times \left(\int \psi_n^{(0)*}(0) \, \hat{V} \psi_1^{(0)}(0) \, dx \right)_{t=0} \quad (32.33)$$

Since the system as a whole is assumed to be closed, its Hamiltonian does not depend upon time explicitly either in the term $\hat{\mathscr{H}}_0$ or in the term $\hat{V}$. Hence, the dependence of any matrix element upon time can be written similar to (32.33):

$$V_{nm} = e^{i\,(E_n - E_m)\,t/h} (V_{nm})_{t=0} \quad (32.34)$$

Assuming that at the initial instant $t = 0$ the system was in a state with energy E_1, that is, in the state $\psi_1^{(0)}$, we must put $C_1 = 1$, $C_{n \neq 1} = 0$. Therefore Eq. (32.32) is integrated thus:

$$C_n(t) = \frac{e^{i\,(E_n - E_1)\,t/h} - 1}{E_n - E_1} (V_{n1})_{t=0}$$

or, again including the exponential factor in the notation of the matrix element, that is, reverting to the functions $\psi_n^{(0)}$ and $\psi_1^{(0)}$, we obtain

$$C_n(t) = \frac{1 - e^{-i\,(E_n - E_1)\,t/h}}{E_n - E_1} V_{n1} \quad (32.35)$$

Consequently, from (25.13) the probability that at time t the system will occur in the state with label n is

$$w_n(t) = |C_n(t)|^2$$
$$= \frac{(1 - e^{-i\,(E_n - E_1)\,t/h})(1 - e^{i\,(E_n - E_1)\,t/h})}{(E_n - E_1)^2} |V_{n1}|^2$$
$$= 2 \left(1 - \cos \frac{(E_n - E_1)\,'t}{h} \right) \frac{|V_{r1}|^2}{(E_n - E_1)^2}$$
$$= 4 \sin^2 \frac{(E_n - E_1)\,t}{2h} \times \frac{|V_{n1}|^2}{(E_n - E_1)^2} \quad (32.36)$$

By definition, the final state belongs to a continuous spectrum, so that we can write simply E instead of E_n. Besides, it is more interesting to find the total probability of the system's passing to a continuous energy-spectrum state. For this we must multiply (32.36) by the number of states belonging to the continuous-spectrum energy interval dE and integrate over the energy. An example of such an expression of the number of states, $dN(E)$, was given in Section 28 (see (28.25)).

We write Eq. (28.25) in general form as follows

$$dN(E) = g(E) \, dE \quad (32.37)$$

The total probability of the required transition is

$$W = \int w(E_1, E) \, dN(E)$$

$$= \int \frac{4 \sin^2 [(E - E_1) t/(2h)]}{(E - E_1)^2} |V(E, E_1)|^2 \, g(E) \, dE \qquad (32.38)$$

For the sake of clarity we shall write the indices E, E_1 of V not as subscripts but in parentheses, like the arguments of a function, which they actually are with respect to V_{EE_1}. Denoting the sine argument by ξ, $(E - E_1) \, t/(2h) \equiv \xi$, and passing to the integration variable ξ, we obtain

$$W = \frac{2t}{h} \int \frac{\sin^2 \xi}{\xi^2} |V(E_1 + 2h\xi/t, \, E_1)|^2 \, g(E_1 + 2h\xi/t) \, d\xi \qquad (32.39)$$

The function $(\sin^2 \xi)/\xi^2$ has its principal maximum at $\xi = 0$; its next maximum is smaller by a factor of twenty. Therefore the main contribution to the integral (32.39) comes from the values of ξ of the order of unity. But then the time t can always be so chosen that $2h\xi/t$ is much smaller than E_1. In other words, in the arguments of the functions V and g we can legitimately replace $E_1 + 2h\xi/t$ simply by E_1 and take these functions outside the integral with respect to ξ.

We have thus demonstrated that if t is of sufficient duration, then the energies of the initial and final states, E_1 and E, are defined so precisely that they can simply be assumed equal, in agreement with the energy conservation law in a transition occurring in a closed system. Of course, the energy conservation law always holds, but for too small values of t it is impossible to define the energy of the final state accurately, since the uncertainty relation (31.37) for this case has the form $(E - E_1) \, t \geqslant 2\pi h$. But when $t \to \infty$, we obtain an exact equality, $E = E_1$.

Since the function $(\sin^2 \xi)/\xi^2$ decreases rapidly with increasing ξ, the integration may be extended from $-\infty$ to ∞. Since the other quantities were taken outside the integral sign, the integral itself is equal to the number

$$\int_{-\infty}^{\infty} \frac{\sin^2 \xi}{\xi^2} \, d\xi = \pi \qquad (32.40)$$

Hence

$$W = \frac{2\pi}{h} |V(E = E_1, \, E_1)|^2 \, g(E_1) \, t \qquad (32.41)$$

Then the probability of a transition in unit time is

$$\frac{dW}{dt} = \frac{2\pi}{h} |V(E = E_1, \, E_1)|^2 \, g(E_1) \qquad (32.42)$$

In this notation it is specifically stressed that $V \, (E = E_1, \, E_1)$ is not a diagonal element of the perturbing potential, but only the matrix element corresponding to the transition of the system to a state with a continuous energy spectrum.

The initial and final states were mutually degenerate, that is, they corresponded to the same energy, and the perturbation "mixed" them.

EXERCISES

1. The potential energy of a system is a homogeneous function of the nth degree in all the coordinates. Find the relation between the integrals

$$\langle T \rangle = - \int \psi^* \frac{h^2}{2m} \nabla^2 \psi \, dV, \qquad \langle U \rangle = \int \psi^* U \psi \, dV$$

that is, between the mean kinetic and mean potential energies of the system.

Solution. Suppose that the length scale has changed α-fold, $r \to \alpha r$. Since ψ remains normalized, the product $\psi^* \psi \, dV$ does not change with the scale. The kinetic energy, which involves the Laplacian, receives the factor α^{-2}, and the potential energy receives, by definition, the factor α^n. Hence the mean energy of the system $\langle \mathscr{H} \rangle = \langle T \rangle + \langle U \rangle$ transforms in the following way:

$$\langle \mathscr{H}_\alpha \rangle = \frac{\langle T \rangle}{\alpha^2} + \alpha^n \langle U \rangle$$

But from the extremality of eigenvalues, the derivative of $\langle \mathscr{H}_\alpha \rangle$ in stationary state must be zero:

$$\frac{\partial \langle \mathscr{H}_\alpha \rangle}{\partial \alpha} = - \frac{2 \langle T \rangle}{\alpha^3} + n \alpha^{n-1} \langle U \rangle = 0$$

Reverting to the initial scale, we put $\alpha = 1$, so that

$$\langle U \rangle = \frac{2}{n} \langle T \rangle$$

In particular, for the Coulomb interaction $n = -1$. Then

$$\langle U \rangle = - 2 \langle T \rangle, \qquad \langle \mathscr{H} \rangle = \frac{1}{2} \langle U \rangle$$

as was obtained in Section 22 on the basis of classical laws.

2. A linear harmonic oscillator is subjected to a perturbation of the form $V = ax^4$. Show that in the first approximation the correction to the nth energy eigenvalue is equal to

$$E_{n1} = \frac{3}{2} a \, \frac{h^2}{m^2 \omega^2} \left(n^2 + n + \frac{1}{2} \right)$$

(make use of the matrix elements (27.28a)).

3. An unperturbed system has two close energy levels the difference between which is comparable with the matrix elements of the perturbing potential between these states. Determine the energy correction in the first approximation.

Solution. By analogy with the perturbation method applied to degenerate states, when the energy eigenvalues of different states are strictly equal, we seek the wave function in the form

$$\psi = C_1 \psi_{01} + C_2 \psi_{02}$$

Here, ψ_{01} and ψ_{02} are the wave functions of both states without the perturbation. This wave function will describe two new states to which the initial states transform as a result of the perturbation. We substitute this function into the equation for finding the eigenvalues of $\hat{\mathscr{H}} = \hat{\mathscr{H}}_0 + \hat{V}$:

$$\hat{\mathscr{H}} \psi = (\hat{\mathscr{H}}_0 + \hat{V}) (C_1 \psi_{01} + C_2 \psi_{02}) = E \psi = E (C_1 \psi_{01} + C_2 \psi_{02})$$

We multiply this equation once by ψ_{01}^* and once by ψ_{02}^*, and integrate over the whole volume. Besides, we make use of the fact that $\hat{\mathscr{H}}_0 \psi_{0i} = E_{i0} \psi_{0i}$, and of the orthogonality and normalization of the eigenfunctions. We obtain a set of two linear homogeneous algebraic equations:

$$(E_{10} + V_{11} - E) C_1 + V_{12} C_2 = 0$$
$$V_{21} C_1 + (E_{20} + V_{22} - E) C_2 = 0$$

It has a solution, provided the determinant developed from the coefficients vanishes:

$$E^2 - (E_{10} + E_{20} + V_{11} + V_{22}) E + (E_{10} + V_{11}) (E_{20} + V_{22}) - |V_{12}|^2 = 0$$

which yields two values for the "perturbed" energy:

$$E = \frac{1}{2} (E_{10} + E_{20} + V_{11} + V_{22})$$
$$\pm \left[\frac{(E_{10} + V_{11} - E_{20} - V_{22})^2}{4} + |V_{12}|^2 \right]^{1/2}$$

If the transformation matrix element V_{12} vanishes, we obtain a conventional formula for first-approximation energy corrections: to E_{10} is added V_{11}, to E_{20} is added V_{22}. From the condition of the problem the matrix element V_{12} is of the same order as $E_{10} - E_{20}$, so that the energy levels are substantially rearranged with respect to the initial configuration.

4. Find the functions ψ^+ and ψ^- of the preceding problem, taking into account that $|C_1|^2 + |C_2|^2 = 1$. Make use of the notation

$$\frac{2V_{12}}{E_{10} + V_{11} - E_{20} - V_{22}} \equiv -\tan \alpha e^{i\beta}$$

Answer.

$$\psi^+ = \cos \frac{\alpha}{2} e^{i\beta/2} \psi_{01} - \sin \frac{\alpha}{2} e^{-i\beta/2} \psi_{02}$$

$$\psi^- = \sin \frac{\alpha}{2} e^{i\beta/2} \psi_{01} + \cos \frac{\alpha}{2} e^{-i\beta/2} \psi_{02}$$

where the $+$ and $-$ signs correspond to the choice of signs in the energy formula.

33

MANY-ELECTRON SYSTEM. THE ATOM

In examining the law of addition of the angular momenta of several particles or the total parity of the state of an atom, we touched upon questions associated with the many-body problem in quantum mechanics. But in quantum mechanics a system consisting of identical particles possesses a very special property stemming from the quantum laws of motion; this property will be examined in the present section. It involves the physical identity of like particles. Since it is impossible to trace the motion of every such identical particle— there is no such thing as path in quantum mechanics—there is no way of indicating the state of a certain selected electron.

In classical mechanics, the state of affairs is rather different: at the initial time one can adopt some convention for numbering identical particles and then, by virtue of the continuity of classical motion, indicate the specific electron at a given point of space moving with a given velocity.

In quantum mechanics such a statement of the problem is physically meaningless. Instead, the state of a system of many identical particles must be defined as follows: list the possible states of an individual particle and indicate the number of particles to be found in each of them. A more detailed definition is incompatible with the fundamental principles of quantum mechanics.

Pauli's Exclusion Principle. In the specific case of a many-electron system, experimental data impose the following additional restriction: not more than one electron can be found in each of the separate states. A state may be either occupied or unoccupied by one electron. This statement is known as *Pauli's exclusion principle*. It is an additional restriction which does not derive from the fundamentals of quantum mechanics we investigated up till now. Its confirmation is to be found in the quantum field theory.

We confine ourselves to the following statement: Pauli's principle is applicable to all particles possessing half-integral spin, that is, electrons, neutrons, and hyperons (hyperons are, so to say, the excited state of a nucleon: they have a greater mass than a nucleon, and half-integral spin); it does not apply to particles with zero or integral spin, among which electromagnetic field quanta may be listed.

With respect to the atom, Pauli's exclusion principle is conveniently formulated as follows: in one and the same atom no two electrons can have the same set of values for the four quantum num-

bers: the principal quantum number n, the orbital quantum number l, the magnetic quantum number k, and the spin quantum number σ. (The spin quantum number is a measure of the spin projection on the same axis on which the orbital angular momentum is projected.)

Sometimes, instead of the four above numbers it is useful to adopt the following: the principal quantum number, the total angular momentum $j = |\, \mathbf{l} + \mathbf{s}\,|$, the orbital quantum number, which in the present case states how the orbital angular momentum is added to the spin (that is, whether they are parallel or antiparallel), and the projection of the total angular momentum on an axis.

According to this choice of quantum numbers, the following system of notation is adopted. First we write the principal quantum number of the electron, n; then the orbital quantum number, but rather than its value we write the corresponding state, s, p, d, or f; the value of the total angular momentum, j, is written as a subscript. Thus, the notation contains three of the four quantum numbers. With respect to the fourth number (the projection k_j) there is a degeneracy. For a given j this fourth number may take $2j + 1$ values. Consequently, according to Pauli's exclusion principle, no more than $2j + 1$ electrons may occur in an nlj state.

The number of electrons in an atom having the given three quantum numbers n, l, and j, is denoted in the form of an exponent attached to the spectroscopic notation of the state taken in parentheses (the number of electrons being read as one would an exponent: square, cube, etc.). For example, if there are two electrons in a state with $n = 2$, $l = 1$, $j = 1/2$, the state as a whole is written $(2p_{1/2})^2$. Obviously, the exponent cannot be greater than $2j + 1$.

Addition of the Angular Momenta of Two Electrons Having the Same n and l. When should the quantum numbers n, l, k, σ be used, and when n, l, j, k_j? In Section 31 it was shown that spin-orbit interaction is due to magnetic forces. But as we shall see later, more important is a special type of interaction between the orbital and spin angular momenta of different electrons, which is of a purely electrostatic origin. In different cases one or another type of interaction, or coupling, may predominate.

If the angular momenta of individual electrons are more strongly coupled, the resultant atomic state is developed as follows: the orbital angular momenta combine, yielding a total orbital angular momentum $\mathbf{L} = |\, \sum \mathbf{l}\,|$; the spin angular momenta combine, yielding a total spin $\mathbf{S} = |\, \sum \mathbf{s}\,|$; and only then the total orbital and total spin angular momenta combine, due to magnetic forces, into the *total angular momentum* $\mathbf{J} = |\, \mathbf{L} + \mathbf{S}\,|$. This type of coupling is called *normal*.

If the spin and orbital angular momenta of each electron combine first, and after that the angular momenta of the individual electrons

combine into the total angular momentum, the coupling is said to be anomalous.

Leaving aside for the time being the question of the causes responsible for one or another type of coupling, we shall examine the rules for the addition of the orbital and spin angular momenta of individual electrons, which differ somewhat from the general rules for the composition of momenta due to Pauli's exclusion principle. Certain additional restrictions appear which are due to the fact that no two electrons can occur in a state with the same four quantum numbers. The following very simple example offers an idea of how Pauli's exclusion principle is applied.

If two electrons have different principal quantum numbers n, or different orbital quantum numbers l, in adding their angular momenta Pauli's exclusion principle can be neglected: at least one of the quantum numbers differs. But if n and l are the same, it should be borne in mind that not all conceivable states of one electron are compatible with the states of the other electron. The same refers to the addition of the angular momenta of a greater number of electrons with the same n and l. Such electrons usually have very close energy values, and the states of atoms are grouped according to them; it is said that the electrons with the same n and l belong to the same *shell* of the atom.

Let us consider the simplest case of $n = 1$. Then, from the definition (29.40), $l = 0$. But at zero l the magnetic quantum number k is also zero. Hence, the electrons have the same three quantum numbers, and according to Pauli's exclusion principle the fourth quantum number σ must necessarily differ. But σ can have only two values, 1 or -1, according to the two spin projections $\pm 1/2$. From Pauli's exclusion principle, each value of σ can belong to one electron with the given n, l, and k, equal respectively to 1, 0, and 0. Hence the resultant state possesses spin $S = 1/2 - 1/2 = 0$. If Pauli's exclusion principle were not taken into account, the total spin could be unity.

Before considering the more general case, we shall introduce a system of notation referring to the shell as a whole. It is analogous to the designation of the states of a separate electron, only instead of lower-case Latin letters we use upper-case letters. The concept of principal quantum number as applied to several electrons is meaningless, and we write only the total orbital angular momentum S, P, D, or F, according to what L is equal to. The term $2S + 1$, where S is the total spin of the electrons, is written as a superior prefix to the symbol; the total angular momentum J is written as a subscript, and the total parity of the electrons is written as a superscript. Even states are called *gerade states* and labeled with the superscript g; odd states are called *ungerade states* and are labeled with the superscript u. If it is necessary to specify in greater detail the

states of the individual electrons that yielded the resultant state, the shell distribution of the electrons according to quantum numbers is attached in the notation described above.

For example, in the case of two electrons with $n = 1$, $l = 0$, $k = 0$, the resultant state of the shell is denoted as:

$$(1s_{1/2})^2 \, {}^1S_0^g$$

Here, only one resultant state is possible. But several resultant states are also possible for the same distribution of individual electrons over different quantum numbers. We shall now examine such a case.

Let there be a state $(np)^2$, that is two electrons having the same principal and orbital quantum numbers, equal to 1. Either their magnetic or their spin quantum numbers must differ, or both. A p-electron can be in one of six states, which we list writing the magnetic quantum number first and the spin projection second:

$$A: 1, \frac{1}{2}; \quad B: 0, \frac{1}{2}; \quad C: -1, \frac{1}{2}; \quad D: 1, -\frac{1}{2}$$

$$E: 0, -\frac{1}{2}; \quad F: -1, -\frac{1}{2}$$

It follows that two electrons can occupy any two different states of the six. As is known, the number of combinations of six things, two at a time, is equal to $C(6, 2) = (6 \times 5)/(1 \times 2) = 15$. These fifteen states differ in the total orbital angular momentum $\mathbf{L}$ and the total spin $\mathbf{S}$, as well as in their projections. The latter depend upon the choice of coordinate axes and will interest us only insofar as they characterize the relative directions of $\mathbf{L}$ and $\mathbf{S}$.

As we know, a state with a given angular momentum is defined by its maximum projection, that is, by the maximum possible value of L compatible with Pauli's exclusion principle, and correspondingly, the largest projection of S. Of the fifteen states, we must in any case select only those for which the total projections, L_z and S_z, are positive or zero, since negative projections cannot, obviously, be the largest.

Eight of the fifteen states have positive projections, and from those eight we select the ones with the largest. We rewrite all eight states compatible with Pauli's exclusion principle:

$$AB: 1, 1; \quad AC: 0, 1; \quad AD: 2, 0; \quad AE: 1, 0$$

$$AF: 0, 0; \quad BD: 1, 0; \quad BE: 0, 0; \quad CD: 0, 0$$

The second number in each pair denotes the spin projection, that is, it involves the necessary factor 1/2.

Now we take the states with the greatest angular momentum projections from among those listed. We start with AB. To it corresponds

an orbital angular momentum projection equal to unity and a spin projection also equal to unity. Each of them may take on zero values (we agreed in advance not to consider negative values). Consequently, we can disregard AC, AE, and AF at once. Of the remaining, let us consider AD. An angular momentum with the maximum projection 2 has positive projections 2, 1, 0; hence, BD and BE should be discarded as possible projections of AD. There remains CD, which has no projections. Obviously, it does not matter whether we take AE or BD for the projection of AB: this does not affect the counting of the number of states.

Let us now write the resultant states in spectroscopic notation, taking into account that the total angular momentum varies from $L + S$ to $|L - S|$:

$$AB:\ (np)^2\,{}^3P_2^g,\quad (np)^2\,{}^3P_1^g,\quad (np)^2\,{}^3P_0^g$$

$$AD:\ (np)^2\,{}^1D_2^g$$

$$CD:\ (np)^2\,{}^1S_0^g$$

Let us also consider the case of three p-electrons. For them we have seven states compatible with Pauli's exclusion principle:

$$ABC:\ 0, \frac{3}{2}\ ;\quad ACE:\ 0, \frac{1}{2};\quad ABD:\ 2, \frac{1}{2}\ ;\quad ABE:\ 1, \frac{1}{2}$$

$$ABF:\ 0, \frac{1}{2}\ ;\quad ACD:\ 1, \frac{1}{2}\ ;\quad BCD:\ 0, \frac{1}{2}$$

The maximum spin projection is 3/2, for a zero orbital angular momentum projection. The maximum orbital angular momentum projection is 2, the total spin projection being 1/2. These two states, together with their projections, are the ones listed above from ABC to ABF. There remains the state ACD, with respect to which BCD can be considered its projection. Thus, the resultant states are

$$ABC:\ (np)^3\,{}^4S_{3/2}^u$$

$$ABD:\ (np)^3\,{}^2D_{5/2}^u,\quad (np)^3\,{}^2D_{3/2}^u$$

$$ACD:\ (np)^3\,{}^2P_{3/2}^u,\quad (np)^3\,{}^2P_{1/2}^u$$

The Wave Equation for a Two-Electron System. Let us now formulate Pauli's exclusion principle with the help of wave functions. To avoid mathematical complications, we shall consider a two-electron system. The wave equation for two electrons should be written as follows:

$$\hat{\mathcal{H}}\Phi = \left(-\frac{h^2}{2m}\nabla_1^2 - \frac{h^2}{2m}\nabla_2^2 + U\,(\mathbf{r}_1, \mathbf{r}_2) \right)\Phi = E\Phi \qquad (33.1)$$

Here, ∇_1^2 and ∇_2^2 are the Laplace operators with respect to the variables of the first and second electrons, and $U\,(\mathbf{r}_1, \mathbf{r}_2)$ is the

potential energy of their interaction with the external field (ext) and between themselves (int):

$$U(\mathbf{r_1}, \mathbf{r_2}) = U_{\text{ext}}(\mathbf{r_1}, \mathbf{r_2}) + U_{\text{int}}(\mathbf{r_1}, \mathbf{r_2}) \tag{33.2}$$

For example, in a helium atom

$$U(\mathbf{r_1}, \mathbf{r_2}) = -\frac{2e^2}{r_1} - \frac{2e^2}{r_2} + \frac{e^2}{|\mathbf{r_1} - \mathbf{r_2}|} \tag{33.3}$$

The wave function depends on the spatial and spin variables of both electrons:

$$\Phi = \Phi(\mathbf{r_1}, s_1; \mathbf{r_2}, s_2) \tag{33.4}$$

The interaction between the spin and orbital angular momenta is weak, at least in the case of low-Z elements (see Sec. 30). Therefore in the potential energy operator we can, in the first approximation, neglect the spin-orbit interaction, which corresponds to (33.3). If the effect of the spin on the orbital motion is small, the probability of a certain value of the spin and the coordinate is equal to the product of the probabilities of both values, and the probability amplitude Φ also separates into a product of amplitudes:

$$\Phi(\mathbf{r_1}, s_1; \mathbf{r_2}, s_2) = \Psi(\mathbf{r_1}, \mathbf{r_2})\chi(s_1, s_2) \tag{33.5}$$

The probability amplitude of the orbital motion satisfies (33.1), provided it does not involve the spin operator. When the system is placed in an external homogeneous magnetic field **H**, to the Hamiltonian is added the operator

$$\hat{U}_{\text{mag}} = \frac{e}{mc}[(\hat{\sigma}_1 \cdot \mathbf{H}) + (\hat{\sigma}_2 \cdot \mathbf{H})] = \beta|\mathbf{H}|(\hat{\sigma}_{1z} + \hat{\sigma}_{2z}) \tag{33.6}$$

where the z axis is taken in the direction of the field (the sign has been changed to "+" because the charge of the electron is negative). The operation of the operator $\hat{\sigma}_{1z} + \hat{\sigma}_{2z}$ on the spin function χ yields simply the total projection of the spin of both electrons, which in the absence of spin-orbit interaction can be treated as an integral of the motion, that is, as a number. This number is simply added to the Hamiltonian, so that the equation for Φ does not alter its form.

Examining the operator $\hat{\mathcal{H}}$ in (33.1), we see that it is completely symmetric in the coordinates of both electrons, that is, it does not change its form if the first electron is called the second, and the second the first:

$$\hat{\mathcal{H}}(\mathbf{r_1}, s_1; \mathbf{r_2}, s_2) = \hat{\mathcal{H}}(\mathbf{r_2}, s_2; \mathbf{r_1}, s_1) \tag{33.7}$$

But (33.1) is a linear equation; hence, if it does not change in the operation (33.7), the wave function can only be multiplied by

a constant number P:

$$\Phi\,(\mathbf{r}_1,\,s_1;\,\mathbf{r}_2,\,s_2) = P\Phi\,(\mathbf{r}_2,\,s_2;\,\mathbf{r}_1,\,s_1) \tag{33.8}$$

Since $\mathbf{r}_1$, s_1 and $\mathbf{r}_2$, s_2 are involved in the same way in all the equations, in (33.8) they can be interchanged, yielding

$$\Phi\,(\mathbf{r}_2,\,s_2;\,\mathbf{r}_1,\,s_1) = P\Phi\,(\mathbf{r}_1,\,s_1;\,\mathbf{r}_2,\,s_2) \tag{33.9}$$

We substitute (33.9) into (33.8) to get

$$\Phi\,(\mathbf{r}_1,\,s_1;\,\mathbf{r}_2,\,s_2) = P^2\Phi\,(\mathbf{r}_1,\,s_1;\,\mathbf{r}_2,\,s_2)$$

or

$$P^2 = 1, \qquad P = \pm 1 \tag{33.10}$$

In this comparatively simple case of two particles the interchange of all coordinates (space and spin) is similar to the symmetry transformation of their corresponding wave functions in reflection (see (29.44)).

We now introduce the *exchange operators* $\hat{P}_r$ and $\hat{P}_s$, which operate only on the electrons' coordinates and spins. Thus

$$\hat{P}_r\Psi\,(\mathbf{r}_1,\,\mathbf{r}_2) = \Psi\,(\mathbf{r}_2,\,\mathbf{r}_1) \tag{33.11}$$

and

$$\hat{P}_s\chi\,(s_1,\,s_2) = \chi\,(s_2,\,s_1) \tag{33.12}$$

If the wave equation is symmetric with respect to the interchange of $\mathbf{r}_1$ and $\mathbf{r}_2$, or of s_1 and s_2, then the eigenvalues of $\hat{P}_r$ and $\hat{P}_s$ are equal to ± 1.

We denote the set of "spatial" quantum numbers of the first electron by the letter n_1 (instead of n_1, l_1, and k_1), and of the second electron by the letter n_2. Then the spatial wave function can be written in greater detail as

$$\Psi = \Psi\,(n_1,\,\mathbf{r}_1;\,n_2,\,\mathbf{r}_2)$$

and from the requirement (33.10) it follows that

$$\hat{P}_r\Psi\,(n_1,\,\mathbf{r}_1;\,n_2,\,\mathbf{r}_2) = \Psi\,(n_1,\,\mathbf{r}_2;\,n_2,\,\mathbf{r}_1) = \pm\,\Psi\,(n_1,\,\mathbf{r}_1;\,n_2,\,\mathbf{r}_2) \tag{33.13}$$

In the case of the upper sign, function (33.13) is said to be *symmetric*, in the case of the lower sign it is *antisymmetric*. The spin function $\chi\,(\sigma_1,\,s_1;\,\sigma_2,\,s_2)$ possesses similar properties with respect to the exchange operator of the spin variables s_1 and s_2.

Now let us consider the requirements Pauli's exclusion principle imposes on the wave function of a two-electron system. We write the function in the form

$$\Phi\,(n_1,\,\sigma_1,\,\mathbf{r}_1,\,s_1;\,n_2,\,\sigma_2,\,\mathbf{r}_2,\,s_2)$$

A total exchange of the spin and spatial variables in this function is due to the operation of the operator $\hat{P}$ equal to

$$\hat{P} = \hat{P}_r \hat{P}_s \tag{33.14}$$

Operating with (33.14) on the function Φ, we obtain

$$\hat{P}\Phi\,(n_1, \sigma_1, \mathbf{r}_1, s_1;\ n_2, \sigma_2, \mathbf{r}_2, s_2)$$
$$= \Phi\,(n_1, \sigma_1, \mathbf{r}_2, s_2;\ n_2, \sigma_2, \mathbf{r}_1, s_1) \tag{33.15}$$

From (33.10), this function is also either symmetric or antisymmetric. But it is immediately apparent that only the antisymmetric function satisfies Pauli's exclusion principle. Indeed, let the states of both electrons be identical, that is, $n_1 = n_2$ and $\sigma_1 = \sigma_2$. Then, if Φ is antisymmetric, we have

$$\hat{P}\Phi\,(n_1, \sigma_1, \mathbf{r}_1, s_1;\ n_1, \sigma_1, \mathbf{r}_2, s_2)$$
$$= \Phi\,(n_1, \sigma_1, \mathbf{r}_2, s_2;\ n_1, \sigma_1, \mathbf{r}_1, s_1)$$
$$= -\Phi\,(n_1, \sigma_1, \mathbf{r}_1, s_1;\ n_1, \sigma_1, \mathbf{r}_2, s_2)$$
$$= \Phi\,(n_1, \sigma_1, \mathbf{r}_1, s_1;\ n_1, \sigma_1, \mathbf{r}_2, s_2) \tag{33.16}$$

By definition, $\hat{P}$ interchanges only the variables $\mathbf{r}$ and s, but not the quantum numbers n and σ. The first equality in (33.16) denotes the result of the operation of $\hat{P}$, the second takes account of the antisymmetry of the wave function, while the third is obtained from the first expression for Φ by means of an exchange of both quadruplets of arguments referring to each electron separately. The possibility of such an exchange for any function is apparent, since it is immaterial which electron is assumed the first and which the second.

The exchange of the first quadruplet $(n_1, \sigma_1, \mathbf{r}_1, s_1)$ with the quadruplet $(n_1, \sigma_1, \mathbf{r}_2, s_2)$ in the last equality in (33.16) is simply meaningless: it makes no difference which arguments we write first, those referring to the first or to the second electron, that is $(n_1, \sigma_1, \mathbf{r}_1, s_1)$ or $(n_1, \sigma_1, \mathbf{r}_2, s_2)$. Thus, the function $\Phi\,(n_1, \sigma_1, \mathbf{r}_1, s_1;\ n_1, \sigma_1, \mathbf{r}_2, s_2)$ is equal to itself with the sign reversed; hence, it is identically zero.

It is apparent that, given the same quantum numbers, only an antisymmetric function possesses this property: a symmetric one would transform into itself. But if the antisymmetric function of two electrons in the same states is identically zero, then the amplitude of the probability of a two-electron system being in such a state is equal to zero for all values of the variables $\mathbf{r}_1$, s_1, $\mathbf{r}_2$, and s_2. Only an antisymmetric function is compatible with Pauli's exclusion principle.

The same holds for the wave function of a many-electron system: it is antisymmetric with respect to a simultaneous interchange of the spatial and spin variables of any pair of electrons. This is the general formulation of Pauli's exclusion principle.

Self-Consistent Field. At the beginning of this section it was pointed out that the state of an atom can be described by stating the number of electrons in a state with given quantum numbers: one electron or none. But what do the quantum numbers themselves denote in a many-electron system?

Every electron is assumed to be in the field of the nucleus and of all the other electrons; such a field is called *self-consistent*. On the basis of such a model of a many-electron atom it is even possible to compute approximately the energy levels of atoms, as was first done by D. R. Hartree.

V. A. Fock substantially improved the self-consistent field method by taking account of electron identity and Pauli's exclusion principle. Fock's method is based on the variation property of energy eigenvalues, that is, the extremal property of the integral (32.16). In this, the wave function of a two-electron system (we restrict ourselves to two electrons) is in the first approximation chosen in the form of the product of the wave functions of each electron separately. Then these functions are so chosen that the integral (32.16) has an extremum, on condition of retaining the normalization of both functions in accordance with (32.17).

In order to satisfy Pauli's exclusion principle at the same time, the initial wave functions should be selected not simply as a product of the functions, $\psi_1(r_1)\psi_2(r_2)$, but as a superposition of the form

$$\Psi = \frac{1}{\sqrt{2}}\left[\psi_1(r_1)\psi_2(r_2) \pm \psi_1(r_2)\psi_2(r_1)\right] \qquad (33.17)$$

(the wave function subscripts 1 and 2 are quantum numbers).

In the case of the upper (plus) sign this coordinate function is multiplied by an antisymmetric spin function χ, and for the lower (minus) sign, by a symmetric spin function. Such functions will be developed later in this section; for the time being we accept the assertion that when the symmetries of the spatial and spin wave functions are opposite, the whole function Φ is antisymmetric.

If the Hamiltonian of a two-electron system does not depend upon the spins, then the expression (32.16) retains only the integration over spatial variables, while summation over the spin variables in every case yields unity. "Memory" of the spin survives in the spatial wave function only in the sign between the terms.

The factor $1/\sqrt{2}$ in (33.17) is selected for purposes of normalization: if ψ_1 and ψ_2 are normalized and orthogonal, then Ψ is also a normalized function, and (32.16), with the wave function (33.17),

can be treated as the mean energy. The normalization of Ψ is verified in the following way:

$$\int |\Psi|^2 \, dV_1 \, dV_2$$

$$= \frac{1}{2} \int |\psi_1|^2 \, dV_1 \int |\psi_2|^2 \, dV_2 + \frac{1}{2} \int |\psi_2|^2 \, dV_1 \int |\psi_1|^2 \, dV_2$$

$$\pm \frac{1}{2} \int \psi_1^* \psi_2 \, dV_1 \int \psi_2^* \psi_1 \, dV_2 \pm \frac{1}{2} \int \psi_2^* \psi_2 \, dV_2 \int \psi_2^* \psi_1 \, dV_1$$

$$= 1 \tag{33.18}$$

We express the two-electron Hamiltonian as follows:

$$\hat{\mathscr{H}} = \hat{\mathscr{H}}'(\mathbf{r}_1) + \hat{\mathscr{H}}'(\mathbf{r}_2) + \hat{V}_{12} \tag{33.19}$$

Here $\hat{\mathscr{H}}'(\mathbf{r}_1)$ and $\hat{\mathscr{H}}'(\mathbf{r}_2)$ are of the same form, but they depend upon the variables of the first and second electron, while $\hat{V}_{12}$ is the interaction Hamiltonian, equal to $e^2 |\mathbf{r}_1 - \mathbf{r}_2|^{-1}$.

Now, substitute the wave function (33.17) into (33.19) and integrate over dV_1 and dV_2. Here we can always redesignate the integration variables, that is, in each term replace $\mathbf{r}_1$ by $\mathbf{r}_2$ and vice versa. As a result each term occurs twice, and the 2 cancels out with the normalization factor 1/2. We thus obtain

$$\langle \hat{\mathscr{H}} \rangle = \int \psi_1^*(\mathbf{r}_1) \, \psi_2^*(\mathbf{r}_2)$$

$$\times [\hat{\mathscr{H}}'(\mathbf{r}_1) + \hat{\mathscr{H}}'(\mathbf{r}_2) + \hat{V}_{12}] \, \psi_1(\mathbf{r}_1) \, \psi_2(\mathbf{r}_2) \, dV_1 \, dV_2$$

$$\pm \int \psi_1^*(\mathbf{r}_1) \, \psi_2^*(\mathbf{r}_2)$$

$$\times [\hat{\mathscr{H}}'(\mathbf{r}_1) + \hat{\mathscr{H}}'(\mathbf{r}_2) + \hat{V}_{12}] \, \psi_1(\mathbf{r}_2) \, \psi(\mathbf{r}_1) \, dV_1 \, dV_2$$

Taking into account that ψ_1 and ψ_2 are orthogonal and normalized,

$$\langle \hat{\mathscr{H}} \rangle = \int \psi_1^* \hat{\mathscr{H}}'(\mathbf{r}) \, \psi_1 \, dV + \int \psi_2^* \hat{\mathscr{H}}'(\mathbf{r}) \, \psi_2 \, dV$$

$$+ \int \psi_1^*(\mathbf{r}_1) \, \psi_2^*(\mathbf{r}_2) \, V_{12} \, \psi_1(\mathbf{r}_1) \, \psi_2(\mathbf{r}_2) \, dV_1 \, dV_2$$

$$\pm \int \psi_1^*(\mathbf{r}_1) \, \psi_2^*(\mathbf{r}_2) \, V_{12} \psi_1(\mathbf{r}_2) \, \psi_2(\mathbf{r}_1) \, dV_1 \, dV_2 \tag{33.20}$$

Let us examine the meaning of the various terms in the mean energy. The first two define the mean energy of the separate electrons in the first and second states. The third term is the energy of interaction between the electrons, since $e\psi_1^*(\mathbf{r}_1) \, \psi_1(\mathbf{r}_1)$ is the charge density of the first electron, $e_2\psi_2^*(\mathbf{r}_2) \, \psi_2(\mathbf{r}_2)$ is the charge density of the second electron, and $e^2 |\psi_1(\mathbf{r}_1)|^2 \, |\psi_2(\mathbf{r}_2)|^2 |\mathbf{r}_1 - \mathbf{r}_2|^{-1} \, dV_1 \, dV_2$ is the interaction energy of two charge elements. This quantity is developed according to the classical law, and its appearance is

obvious. The last term is of a quantum nature: it appears as a consequence of symmetrization of the wave function (33.17). It is called the *exchange integral*, or *exchange energy*, of the two electrons. The term "exchange" was adopted because the electron is as it were simultaneously in both states.

The sign of the exchange integral depends upon the spin state of the system. Therefore, even though we neglect the magnetic interaction of the spins, it is necessary to take into account their interaction through the exchange integral. Here, the interaction cannot be reduced to some "force": it is due to purely quantum properties of wave function symmetry stemming from Pauli's exclusion principle. The exchange interaction is greater than the magnetic interaction, because it is, in the final analysis, due to the Coulomb potential energy of two electrons, $e^2 \,|\, \mathbf{r}_1 - \mathbf{r}_2 \,|^{-1}$, and does not contain c^2 in the denominator.

We now vary the integrals (33.20) with respect to ψ_1^* and ψ_2^*, provided the functions are orthogonal, and additionally $\int \psi_1^* \psi_1 \, dV = 1$ and $\int \psi_2^* \psi_2 \, dV = 1$. As usual, the variations of these two expressions are multiplied respectively by the parameters $-E_1$ and $-E_2$, and the variations of the orthogonality conditions, $\int \psi_1^* \psi_2 \, dV = 0$ and $\int \psi_2^* \psi_1 \, dV = 0$, by two other parameters, $-\eta_1$ and $-\eta_2$. Denoting the variation with respect to ψ_1^* as $\delta_{\psi_1^*}$, and with respect to ψ_2^* as $\delta_{\psi_2^*}$, yields the extremal condition for $\langle \mathcal{H} \rangle$ for the additional normalization conditions:

$$\delta_{\psi_1^*} \langle \mathcal{H} \rangle - E_1 \delta \psi_1^* \psi_1 - \eta_1 \delta \psi_1^* \psi_2 = 0 \tag{33.21a}$$

$$\delta_{\psi_2^*} \langle \mathcal{H} \rangle - E_2 \delta \psi_2^* \psi_2 - \eta_2 \delta \psi_2^* \psi_1 = 0 \tag{33.21b}$$

After performing the variation we arrive at a system of two integro-differential equations:

$$\hat{\mathcal{H}}'(\mathbf{r}) \psi_1(\mathbf{r}) + e^2 \int \frac{|\psi_2(\mathbf{r}')|^2 \, dV'}{|\mathbf{r} - \mathbf{r}'|} \psi_1(\mathbf{r})$$

$$\pm e^2 \int \frac{\psi_2^*(\mathbf{r}') \psi_1(\mathbf{r}') \, dV'}{|\mathbf{r} - \mathbf{r}'|} \psi_1(\mathbf{r}) = E_1 \psi_1(\mathbf{r}) + \eta_1 \psi_2(\mathbf{r}) \tag{33.22a}$$

$$\hat{\mathcal{H}}'(\mathbf{r}) \psi_2(\mathbf{r}) + e^2 \int \frac{|\psi_1(\mathbf{r}')|^2 \, dV'}{|\mathbf{r} - \mathbf{r}'|} \psi_2(\mathbf{r})$$

$$\pm e^2 \int \frac{\psi_1^*(\mathbf{r}') \psi_2(\mathbf{r}') \, dV'}{|\mathbf{r} - \mathbf{r}'|} \psi_2(\mathbf{r}) = E_2 \psi_2(\mathbf{r}) + \eta_2 \psi_1(\mathbf{r}) \tag{33.22b}$$

Multiplying (33.22*a*) by $\psi_1^*(\mathbf{r})$ and by $\psi_2^*(\mathbf{r})$, and integrating over V, we obtain the expressions for E_1 and η_1:

$$E_1 = \int \psi_1^*(\mathbf{r})\,\mathscr{H}'(\mathbf{r})\,\psi_1(\mathbf{r})\,dV$$

$$+ e^2 \int \frac{|\psi_1(\mathbf{r})|^2 |\psi_2(\mathbf{r}')|^2}{|\mathbf{r}-\mathbf{r}'|}\,dV\,dV'$$

$$\pm e^2 \int \frac{\psi_1^*(\mathbf{r}')\,\psi_2(\mathbf{r})\,\psi_1^*(\mathbf{r})\,\psi_2(\mathbf{r}')\,dV\,dV'}{|\mathbf{r}-\mathbf{r}'|} \quad (33.23)$$

$$\eta_1 = \int \psi_2^*(\mathbf{r})\,\mathscr{H}'(\mathbf{r})\,\psi_1(\mathbf{r})\,dV$$

$$+ e^2 \int \frac{|\psi_2(\mathbf{r}')|^2\,\psi_2^*(\mathbf{r})\,\psi_1(\mathbf{r})\,dV\,dV'}{|\mathbf{r}-\mathbf{r}'|}$$

$$\pm e^2 \int \frac{|\psi_2(\mathbf{r}')|^2\,\psi_2(\mathbf{r})\,\psi_1(\mathbf{r})\,dV\,dV'}{|\mathbf{r}-\mathbf{r}'|} \quad (33.24)$$

Similar expressions are developed for E_2 and η_2.

In the solutions it is always possible to separate out the angular dependence of the wave functions, so that only the equations involving the dependence upon r remain. The final set of equations can be solved by computers, whereas the exact Schrödinger equation for the two-electron problem so far defies solution by any means. The degree to which the solutions obtained from Fock's equations agree with experimental data is in a number of cases quite satisfactory. As always, in employing the variational method, the energy eigenvalues agree with experience better than any integral expressions obtained with the help of the wave functions determined together with the energy.

The self-consistent field method of representing the wave functions offers an understanding of the reason why, in adding the angular momenta of several electrons, some values of the total angular momentum correspond to smaller total energies than others.

Let us take as an example two *p*-electrons. The angular parts of the wave functions of the *p* state are first-order spherical functions, that is Y_1^1, Y_1^0, Y_1^{-1}:

$$Y_1^0 = \cos\vartheta, \qquad Y_1^{\pm1} = \pm\sin\vartheta\,e^{\pm i\varphi}$$

The electron density distribution corresponding to these functions is

$$|Y_1^0|^2 = \cos^2\vartheta, \qquad |Y_1^1|^2 = |Y_1^{-1}|^2$$

In other words, for Y_1^0 the distribution is elongated along the polar axis, while for $Y_1^{\pm1}$ it is flattened in the "equatorial" plane.[10]

---

[10] Figuratively speaking, to nonzero angular momentum projections correspond greater "arms".

The electrons' Coulomb repulsion energy is, as we have just seen, expressed by the integral

$$e^2 \int \frac{|\psi_1(\mathbf{r}_1)|^2 |\psi_2(\mathbf{r}_2)|^2 \, dV_1 \, dV_2}{|\mathbf{r}_1 - \mathbf{r}_2|}$$

If two electrons are in the same orbital states (and consequently have opposite spins), their repulsion energy is greatest, because they are moving in one spatial domain; if the orbital electrons are in different orbital states, the energy is smaller, since the electrons are moving at a greater distance from one another.

If we now consider states with magnetic quantum numbers ± 1, that is Y_1^1 and Y_1^{-1}, from the Coulomb repulsion formula we should expect them to display the same interaction energies between an electron pair with angular momentum projections $+1$ as between a pair with projections $+1$ and -1, since the same electron density corresponds to ± 1. Actually, however, there is a difference, due to the exchange term, though it is nevertheless smaller than in comparing an electron pair having the same or opposite projections with a pair having projections equal to 1 and 0, that is $Y_1^1 Y_1^{-1}$ with $Y_1^1 Y_1^0$.

Consequently, in adding the angular momenta of separate electrons, the least-energy state is achieved when the spatial wave functions have the smallest overlap. If the squares of their moduli are the same, as in Y_1^1 and Y_1^{-1}, the exchange effect alters due to the overlap, but it remains greater than for functions with 0 and 1 angular momentum projections.

Therefore, when the angular momenta of two p-electrons are added their spin must have the maximum value, so that the projections of orbital angular momentum would correspond to the least overlapping wave functions, Y_1^1 and Y_1^0. In consequence, we find that a system of two p-electrons has the largest spin, 1, and for this greatest spin the greatest total orbital angular momentum. We have arrived at what is known as *Hund's first rule*.

If we have three p-electrons, the orbital angular momentum of the third must already have the magnetic quantum number -1, and the total spin is then $3/2$. From the energy point of view this is more advantageous than if the spin of the third electron was in the opposite sense of the first two electrons, which would be in the same orbital state. Hence, of the three possible resultant states of a system, $(np)^3$, found above, the $^4S_{3/2}$ state has the least energy. On the other hand, the states $^2D_{3/2\ 5/2}$ and $^2P_{1/2\ 3/2}$ have two coinciding projections of the electrons' orbital angular momenta: for state $^2D_{3/2\ 5/2}$ this yields two wave functions Y_1^1, for state $^2P_{1/2\ 3/2}$, two functions Y_1^0. This confirms Hund's rule: the spin must be put as large as possible, and then for the given spin we seek the maximum possible value of the orbital angular momentum.

The example cited above refers to the nitrogen atom, which has the configuration $(np)^3$. The least-energy state is $^4S_{3/2}$, as expected; the $^2D_{3/2\ 5/2}$ states lie 2.2 eV higher, and the $^2P_{1/2\ 3/2}$ states lie 3.8 eV higher. The latter are highest because when two electrons are moving close to the polar axis in a Y_1^0 state, they are on the whole at a smaller distance than when they move close to the equatorial plane. This can be easily seen by constructing polar diagrams of $\sin^2\vartheta$ and $\cos^2\vartheta$ and rotating them about the polar axis.

This provides an explanation of normal coupling of the angular momenta of separate electrons. The coupling is due to the Coulomb repulsion and to Pauli's exclusion principle. In atoms that are not too heavy, normal coupling due to Coulomb forces always exceeds the spin-orbit coupling, which is due to magnetic forces.

This is the way orbital and spin angular momenta of separate electrons are added. However, between the total orbital and total spin angular momenta of a system of electrons there exists a magnetic interaction similar to that found for an individual electron.

Unlike a separate electron, the spin of a system of several electrons may be greater than 1/2: by Hund's first rule, larger values of spin are preferred. Therefore, in the most general case the total angular momentum J has, instead of two possible values, as many values as there are values of the vector sum $J = |\ \mathbf{L} + \mathbf{S}\ |$. If $L > S$, the sum has $2S + 1$ values, if $L < S$, it has $2L + 1$ values.

Multiplets. If there were no spin-orbit magnetic interaction, each term could freely orient by itself. This would result in a $(2S + 1) \times \times (2L + 1)$-fold degeneracy. Magnetic interaction partially lifts the degeneracy: when the orbital angular-momentum vector is fixed, a magnetic field is as it were created; it possesses axial symmetry for which to each spin projection there corresponds a specific spin-orbit interaction energy. (If the spin is larger a similar reasoning applies to the orbital angular momentum.)

If the projection of the smaller angular momentum on the larger is given, then, apparently, the total angular momentum is determined according to the general rule for adding angular momenta. The total angular momentum again can rotate freely in space, but with $(2J + 1)$-fold degeneracy. Thus, the magnetic interaction splits $(2S + 1)(2L + 1)$ levels into $2S + 1$ or $2L + 1$ levels with the given value of the total angular momentum J, each level with definite J being $(2J + 1)$-fold degenerate. The totality of all these levels is called a *multiplet*, and the levels themselves are known as *fine-structure levels*.

A multiplet has $2S + 1$ or $2L + 1$ components. One of them corresponds to the least energy. It is determined as follows. For a given orbital quantum number l there may be $2l + 1$ values of the magnetic quantum number, and for a given magnetic quantum

number there are two spin projections. Hence, there can be $2(2l + 1)$ electrons in the given configuration. It can readily be observed that if there are $2l + 1$ electrons in it, the total orbital angular momentum should be zero. Then the wave functions of all the separate electrons are different, and the Coulomb repulsion energy is least. This was examined in greater detail in the example on the $(np)^3$ configuration.

Thus, filling of the $2 (2l + 1)$ states takes place as it were in two stages: first $2l + 1$ states with the total angular momentum $L = 0$ are filled, and only then the other $2l + 1$ states. We find that when the former set of states are being filled, the least multiplet level corresponds to the value $J = |L - S|$, while in filling the subsequent $2l + 1$ states the least level has $J = L + S$. This is known as *Hund's second rule*; it will be substantiated in considering the fine structure of atomic levels.

Let us also determine how many electrons in an atom can have a given principal quantum number n. Since l varies from zero to $n - 1$, and since for a given n and l there can be $2 (2l + 1)$ electrons, we find that in an atom Pauli's exclusion principle permits a total of

$$\sum_{l}^{n-1} 2(2l+1) = 2n^2 \qquad (33.25)$$

electrons with the given n.

The Thomas-Fermi Method. Fock's method gives satisfactory quantitative results when applied to the problem on the motion of several electrons. But for a large number of electrons the equations, naturally, become so complex and involved that it is difficult to derive any general laws from them. For such cases there is a more approximate, but very general, method of approach to the many-electron problem based precisely on the fact that the number of electrons in the atom, Z, is great in comparison with unity.

This approximate method was suggested independently by Thomas and Fermi on the basis of intuitive, but extremely graphic, reasoning. Subsequently Dirac showed that the Thomas-Fermi approximation could be developed from Fock's equations by applying the quasiclassical approximation. The precision of the method, that is, the permissible error, is of the order $Z^{-2/3}$.

Despite the fact that it is always preferable to present the stricter method of deduction, enabling an evaluation of the error, we shall take a graphic approach to the Thomas-Fermi equations, which offers a better understanding of their physical essence. Furthermore, even if the error in passing from Fock's equations to the Thomas-Fermi equations can be seen, the degree of approximation of Fock's

equations in comparison with the exact quantum mechanical equations has to this day not been evaluated at all. All we know is that the numerical expression of the error is usually not great, but why this is so remains theoretically unexplained.

We shall proceed from the equations of Section 28 for the possible number of particle states. From Eq. (28.23), the number of states of a quantum mechanical particle with linear momenta lying between p_x and $p_x + dp_x$, p_y and $p_y + dp_y$, p_z and $p_z + dp_z$, in a volume V is

$$dN\,(p_x,\ p_y,\ p_z) = \frac{V\,dp_x\,dp_y\,dp_z}{(2\pi h)^3}$$

Since an electron additionally possesses an internal degree of freedom (spin), the right-hand side of this equation should be multiplied by two. Furthermore, electrons are subject to Pauli's exclusion principle whereby no two electrons can be found in each state with a given spin projection.

Let there be a system of N electrons. What is the least possible value of their kinetic energy?

It is obvious that, according to Pauli's exclusion principle, they cannot all have zero kinetic energy, since they would then all have to be in a state with $p_x = p_y = p_z = 0$, which is prohibited. Only two electrons with spin projections of opposite sign can have the same linear momentum (including zero). The next two electrons must have a momentum differing somewhat from zero.

In the ground state for all the electrons as a whole each pair of electrons must occupy a vacant state lying as close as possible to the zero-momentum state and unoccupied by another pair. It is obvious that a sufficiently large number of electrons will occupy a sphere in momentum space centred at the origin of the coordinate system. The volume of this sphere is made up of separate cubes of volume $(2\pi h)^3/V$, each cube containing two electrons. Any deviation from the spherical shape of the volume in a momentum space filled with electrons leads to an increase in the total energy, that is, a deviation from the ground state.

Let the greatest electron energy within this sphere be E_0. Then from Eq. (28.25) it is easy to relate E_0 to the total number of electrons in the sphere. Adding the factor 2, which takes account of spin, we obtain

$$N = \frac{2^{1/2} m^{3/2}}{\pi^2 h^3}\, V \int\limits_0^{E_0} E^{1/2}\, dE = \frac{2^{3/2} m^{3/2} V}{3\pi^2 h^3}\, E_0^{3/2}$$

The ratio N/V is essentially the electron density n, so that the relation between the boundary electron energy E_0 and the density is

$$E_0 = 3^{2/3}\pi^{4/3}\, \frac{h^2}{2m}\, n^{2/3} \tag{33.26}$$

Actually, this relation refers to the kinetic rather than the total energy of the electrons. The difference between the total energy and the kinetic energy is essential only when the motion occurs in a domain in which the potential energy varies.

Let us consider the potential-energy curve in Figure 38. The kinetic energy of a particle in the domain $0 \leqslant x \leqslant a$ is plotted from $U = 0$, and in the domain $x > a$, from $U = U_0$. (We assume that the domains are large enough to apply the quasi-classical approximation for the given kinetic energy, that is, that they are
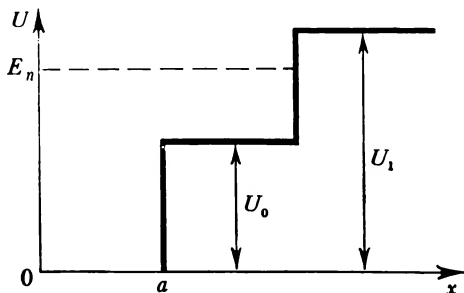
Figure 38

substantially greater than the de Broglie wavelength. In this case the kinetic and potential energies have approximate meaning as separate quantities.)

For such a potential-energy curve the least total energy is obtained when its maximum value $E_0 + U$ is the same in both domains (analogous to the way a liquid in communicating vessels assumes the same level in both). If the maximum electron energy were greater in one domain than in the other, the total energy of the electrons could decrease on account of their passing to vacant places in the domain with lower maximum energy. But we are looking for the ground state of a system, in which the total energy can decrease no more.

Thus, the condition of the electrons being in the state with the least total energy is

$$E_0 + U = \text{constant} \tag{33.27}$$

This condition can also be applied in cases when the potential energy varies spatially not in jumps, as in Figure 38, but smoothly, as in the atom. It is only necessary for it not to vary too greatly over a distance equal to one de Broglie electron wavelength, in accordance with the general provision for the applicability of the quasi-classical approximation (see Sec. 31). But when the number of

electrons is large their maximum kinetic energy is great, hence the de Broglie wavelength is small. Therefore, the condition for the application of this method is that the number of electrons in an atom be sufficiently large in comparison with unity.

The potential energy distribution in the atom looks approximately like the curve presented in Figure 39. The potential energy is negative everywhere because it is gauged to zero at infinity.

The boundary energy of the electrons should not be positive anywhere, because with positive energy they could recede from the
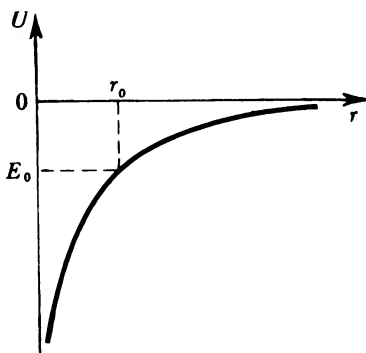
Figure 39

atom into infinity. Hence, the boundary energy can be either negative or zero. We shall show that it cannot be negative, that is, it is zero. If, for example, the boundary energy were defined by the dashed line in Figure 39, then the electron density would have to vanish at the point $r = r_0$: from Eq. (33.26) at the point where the kinetic energy is zero the density also vanishes.

If we assume that at some point $r = r_0$ the electron density is zero, then we must accept that all the electrons are to be found at $r \leqslant r_0$, so that the total charge of the atom, both the positive and the negative, is concentrated in a sphere of radius $r = r_0$. But then, by the Gauss law, the electric field must become zero at point $r = r_0$, since the charge distribution is taken to be spherically symmetrical. But $|\mathbf{E}| = -d\varphi/dr$, and from Figure 39 it is apparent that the derivative of the potential at this point is not zero. Consequently, the only possibility is that the whole charge is concentrated in a domain on whose boundary the derivative of the potential vanishes.

In Figure 39, the derivative of the potential tends to zero at $r \to \infty$. It follows that the boundary energy of the electrons is equal to zero, as asserted. However, it is possible to visualize a case when the potential energy curve approaches the abscissa at a horizontal tangent for finite r. This also yields zero for the boundary energy.

It will be shown further on, however, that this does not occur. In any case, we shall proceed from the consideration that the boundary energy is equal to zero. The constant in Eq. (33.27) is precisely that boundary energy.

The potential energy of an electron is $U = -e\varphi$, while the kinetic energy must be expressed in accordance with (33.26). We then obtain the basic relationship for the Thomas-Fermi method between the potential in the atom and the electron density at a given point:

$$3^{2/3}\,\pi^{4/3}\,h^2\,(2m)^{-1}\,n^{2/3} - e\varphi = 0 \tag{33.28}$$

We obtain the second relation between the potential and the density from Eq. (16.7). We put the opposite sign in the right-hand side since the charge of an electron is negative:

$$\frac{1}{r^2}\frac{d}{dr}r^2\frac{d\varphi}{dr} = 4\pi en \tag{33.29}$$

Solving Eq. (33.28) for the electron density and substituting it into (33.29), we obtain an equation describing the electron distribution in the atom:

$$\frac{1}{r^2}\frac{d}{dr}r^2\frac{d\varphi}{dr} = \frac{2^{7/2}}{3\pi}\frac{m^{3/2}}{h^3}e^{5/2}\,\varphi^{3/2} \tag{33.30}$$

We transform this equation in the same way as (20.6). For this we substitute φ in the form

$$\varphi = \frac{Ze}{r}\,\psi \tag{33.31}$$

The function ψ is dimensionless, since Ze/r has the dimensions of potential. In the immediate vicinity of the nucleus, φ is determined only by the nucleus, since its potential tends to infinity like Ze/r, while the potential of a spatially distributed electron charge remains finite. Hence, close to the nucleus the boundary condition consists in that $\psi(0) = 1$.

At large distances from the nucleus its charge is completely screened by the charge of the electrons. Therefore the potential must tend to zero faster than $1/r$. Hence $\psi(\infty) = 0$.

Substituting (33.31) into (33.30), we find the equation for ψ:

$$\frac{d^2\psi}{dr^2} = \frac{2^{7/2}}{3\pi}Z^{1/2}\frac{m^{3/2}}{h^3}e^3\,\frac{\psi^{3/2}}{r^{1/2}} \tag{33.32}$$

It is convenient to eliminate the dimensional factor in the right-hand side. For that we have to introduce a new unit of length similar to the atomic unit (see (29.30)):

$$r = \frac{(3\pi)^{2/3}}{2^{7/3}}\frac{1}{Z^{1/3}}\frac{h^2}{me^2}\,x \tag{33.33}$$

This unit differs from the atomic unit by the factor $0.889/Z^{1/3}$. Introduction of the dimensionless variable x reduces (33.30) to the standard form

$$\frac{d^2\psi}{dx^2} = \frac{\psi^{3/2}}{x^{1/2}} \tag{33.34}$$

of the *Thomas-Fermi equation*.

Now neither the equation nor the boundary conditions for it, that is $\psi(0) = 1$ and $\psi(\infty) = 0$, involve the atomic number. It is therefore sufficient to integrate (33.34) once for all atoms. But the equation can, of course, be applied only to atoms with large and medium atomic numbers.

Going back to the dimensional radius r, we find that the function $\psi(x)$ yields the potential distribution for each Z:

$$\varphi = \frac{Ze}{r}\, \psi\left(1.125 Z^{1/3}\, \frac{me^2}{h^2}\, r\right) \tag{33.35}$$

If the distance from the nucleus is expressed in terms of x, the electron density distribution appears the same for all atoms. This, naturally, is a property of Eq. (33.34), and not of real atoms. But even such a highly approximate treatment makes it possible to draw important conclusions concerning the atom as a whole.

Since for the same values of x the value of $\psi(x)$ is the same for all atoms, the corresponding values of r for different atoms are inversely proportional to $Z^{1/3}$. Hence, in heavy atoms the bulk of the electrons is concentrated closer to the nucleus than in lighter atoms.

We shall now show that the boundary condition $\psi = 0$ can be imposed in a noncontradictory manner only for $x = \infty$. It was already mentioned that the potential curve must approach the point x_0 where $\psi(x)$ vanishes only with a horizontal tangent. Consequently, close to that point the expansion of $\psi(x)$ commences with such a term:

$$\psi(x) = a\,(x - x_0)^{2+k}$$

where k is a positive number. Substituting this expansion into Eq. (33.34), we have

$$(2 + k)\,(1 + k)\, a\,(x - x_0)^k = \frac{a^{3/2}}{x_0^{1/2}}\,(x - x_0)^{3+3k/2}$$

whence it follows that $k = -6$, contrary to the requirement that $k \geqslant 0$. Only an asymptotic osculation with the abscissa does not lead to a contradiction.

The Appearance in Atoms of Electrons with a Given Value of *l*. In developing the Thomas-Fermi equation we proceeded from a

momentum distribution of electrons. However, the question can also be posed of their distribution according to other integrals of the motion. Since in the atom, that is, in a central field, angular momentum is conserved, it is natural to seek an electron distribution according to angular momenta.

Let us first see what the angular momentum distribution of the electrons in an atom is like. The boundary momentum of the electrons is proportional to the square root of the boundary kinetic energy E_0, since from (33.26) the electrons having the largest momenta lie closer to the nucleus. But the angular momentum is proportional to the product of the linear momentum multiplied by the distance from the nucleus, hence close to the nucleus an electron's angular momentum is small. At greater distances from the nucleus the electron density decreases, and accordingly the boundary momentum decreases too. This again leads to a decrease in angular momentum. It follows, then, that the angular momentum attains its maximum value somewhere at median distances from the nucleus. This maximum angular momentum is the greater the higher the electron density. That is why in heavy atoms, where electron density is great, angular momentum attains great values.

To determine the maximum values of angular momentum that are possible for a given Z, we shall proceed from the classical expression for the energy of a particle in a central field:

$$E = \frac{p_r^2}{2m} + \frac{M^2}{2mr^2} - \frac{Ze^2\psi}{r} \tag{33.36}$$

In accordance with the basic assumption (33.28), the boundary energy E should be put equal to zero. Then for the radial component of the linear momentum we obtain the following formula:

$$p_r = \left(\frac{2mZe^2\psi}{r} - \frac{M^2}{r^2} \right)^{1/2} \tag{33.37}$$

In place of M^2 we can put $h^2 l\,(l+1)$. But since the present theory corresponds wholly to the quasi-classical approximation, a rather better result is obtained for the quasi-classical eigenvalue of M^2. It can be shown (though we shall not do this) that the quasi-classical eigenvalue of M^2 is $h^2\,(l+1/2)^2$. Note that $(l+1/2)^2$ differs from $l\,(l+1)$ by only 1/4.

We now take h/r outside the parentheses in (33.37) and express the remaining expression in terms of the dimensionless quantity x:

$$p_r| = \frac{h}{r} \left(1.778 Z^{2/3} x\psi - \left(l + \frac{1}{2} \right)^2 \right)^{1/2} \tag{33.38}$$

For p_r to be a real quantity, the radicand must remain real within a certain interval of values of x. But since $x\psi = 0$ at $x = 0$ and at $x = \infty$, the interval is finite and includes the maximum of $x\psi$.

The maximum is equal to 0.488, hence the whole interval in which p_r is a real quantity contracts into a point at a value of Z such that

$$1.778 \times 0.488Z^{2/3} = \left(l + \tfrac{1}{2}\right)^2 \tag{33.39}$$

Then the curve $y = 1.778Z^{2/3}x\psi$ touches the horizontal line $y = (l + 1/2)^2$. It follows that a given value of l in an atom can appear for Z satisfying the condition

$$Z = 0.155\,(2l + 1)^2 \tag{33.40}$$

According to this equation, electrons having $l = 2$ appear at $Z = 19$, and those having $l = 3$, at $Z = 53$. A better agreement with reality is obtained if in the latter formula we take 0.17 instead of 0.155.

The whole of the foregoing calculation is possible only because account is taken of the screening of the nuclear field by the electrons, with which the appearance of the maximum of the function $x\psi$ is associated. Nothing of the kind could occur in a pure Coulomb field. This points to the fact that, as the atomic number increases, so does the dependence of electron energy on the orbital quantum number.

When the field differs substantially from a Coulomb field, the dependence of the energy upon l becomes so strong that an increase in the principal quantum number n, with a simultaneous reduction in l, leads to a slower increase in energy than an increase in n for a given l.

The reason is that for large l the angular momentum arm is large, that is, the electron is far from the nucleus. But in an atom the dependence of an electron's potential energy upon r does not obey the Coulomb law: it decreases much faster, owing to screening. That is why at large distances from the nucleus the electron is as it were ejected from the potential well by a centrifugal force, which leads to a comparatively large rise in the energy level.

The increase in energy in a transfer from the energy level $E\,(n,\,l)$ to $E\,(n + 1,\,0)$ turns out to be less than in a transfer from $E\,(n,\,l)$ to $E\,(n,\,l + 1)$. In a Coulomb field the energy depends only upon n, and a transfer from $E\,(n,\,l)$ to $E\,(n,\,l + 1)$ does not change the energy at all, while a transfer from $E\,(n,\,l)$ to $E\,(n + 1,\,0)$ leads to a rise in level. The inequality $E\,(n + 1,\,0) - E\,(n,\,l) > E\,(n,\,l + 1) - E\,(n,\,l)$ also occurs in a weak deflection of the field from the Coulomb field.

The Mendeleyev Periodic Table. Let us now consider, in general terms, the filling of the spaces in the Mendeleyev Periodic Table. With few electrons in the atoms, the dependence of an electron's energy on the quantum numbers does not differ substantially from that found in a purely Coulomb field, so that the ordering of energy

levels is determined by the principal quantum number n. Accordingly, filling of the atom's electron shells takes place according to increasing values of n.

As can be seen from Eq. (33.25), the shell with $n = 1$ can contain two electrons. *Hydrogen* has one electron, its state is 1s; in *helium* the shell is filled, the state being $(1s)^{21} S_0^g$. The electron shell of the ground state of a helium atom is so stable that if any other atom approaches it the total energy can only increase, so that repulsive forces appear. Helium is chemically inert. The interaction forces between helium atoms are small as a result of the symmetry and stability of their electron shells. Therefore, helium is liquefied at an extremely low temperature.[11]

After helium, the filling of the shell with $n = 2$ begins. The first electron of this shell, that is, a 2s-electron, appears in *lithium*. The two inner 1s-electrons occurring in the helium configuration form a closed shell (the K shell) and strongly screen the nuclear charge; consequently, the outer electron is weakly bound. Such is the alkalimetal electron configuration in the case of lithium, and analogous electron configurations subsequently result each time (Na, K, Rb, Cs) from the addition of an s-electron to a nucleus surrounded by a noble-gas electron shell.

Lithium is followed by *beryllium*, the atom of which has two 2s-electrons, the configuration being $(1s)^2 (2s)^2 {}^1S_0^g$. Although the $(2s)^2$ shell (the L_I shell) is filled, its properties are quite unlike the properties of the noble gas helium with the filled $(1s)^2$ shell. The reason for this is that the field acting on an electron in the light element beryllium still approximates the Coulomb field sufficiently, and the energy of the 2s state is therefore close to the energy of the 2p state: it is but slightly dependent upon the orbital quantum number. A small energy is needed for the transition of an electron from the 1s shell to a 2s or 2p shell. Because of this the electron configuration of beryllium is unstable with respect to perturbations. The small energy that evolves in the joining of other atoms is sufficient to cause the restructuring of the 2p shell required to form a stable system of bound atoms.

After beryllium, filling of the 2p shell (the L_{II} shell) begins, until it is completely filled for *neon*. Neon is preceded by *fluorine*, which lacks one electron to completely fill the L_{II} shell. The energy required to add an electron to the $(2p)^5$ shell of fluorine, completing it as a neon shell, is large. This explains the chemical activity of

---

[11] The condensation of helium into a liquid at low temperatures is due to the so-called Van der Waals' forces, which arise out of the mutual electrostatic polarization of approaching atoms. These forces act at larger distances than the forces of chemical affinity, and are very small compared with them.

fluorine and the other halogens, which are similarly situated with respect to the noble gases.

A shell with $n = 2$ may contain eight electrons. The elements in which the electrons occupy this shell form the second period of the Mendeleyev Periodic Table (H and He form the first period). Then the shell with $n = 3$ is filled, but initially only its first two subshells, 3s and 3p (M_I and M_{II}). The structure of the outer electron shells in elements of the third period is similar to that of the shells in elements of the second period.

The chemical properties of atoms are in the main determined by their outer electron shells. This explains the similarity of chemical properties on which the periodic law is based.

In *argon*, the 3p shell is filled, thus completing one more period of eight elements in the Periodic Table. Argon owes its noble gas configuration to the fact that the 3p state, on the one hand, and the 3d and 4s states, on the other, differ considerably in energy. One could say that in argon there takes place an equalization of the effects of the principal quantum number and the orbital quantum number on the energy level: their effect is approximately the same and still strong enough for the element to have the stable electron shell of a noble gas. But unlike helium, argon is nevertheless capable of forming chemical compounds.

In investigating the states of shells which require less than half the total number of electrons to be filled (that is, less than $2l + 1$), we may consider that the unoccupied states ("holes") behave like electrons. For example, if an np shell lacks two electrons to six, we may combine the states of the two "holes" in the same way as we combine the states of two np-electrons. The result is always correct, but here, to find the total angular momentum of such a system of two "holes", we apply Hund's second rule, that is, take $J = L + S$. By adding the spins and orbital angular momenta, with due account for Pauli's exclusion principle, we can readily verify that the four electrons in the np shell are equivalent to the two "holes" in it (see Exercise 2).

The concept of a "hole" in a system of occupied states and the equivalence of a "hole" and a particle has proved to be extremely useful in many branches of physics where many-electron systems are studied.

Let us now give, in one table, the scheme for building up the first eighteen places in the Periodic Table. It also gives the number of electrons having given quantum numbers.

After argon, the stronger effect of the orbital quantum number than the principal quantum number on the energy comes into play. In other words, from the energy point of view filling of the 4s shell turns out to be preferable to filling of the 3d shell. The new period begins with the alkali metal *potassium*. We note an empirical regu-

larity: filling occurs for the same $n + l$. The sum $n + l$ is the same for the $3p$ and $4s$ shells and is equal to 4, while it is already greater by unity in the $3d$ shell. The $4p$ shell is filled after the $3d$ shell, with the same value of the sum $n + l = 5$, and then the $5s$ shell. It is seen that this rule is observed later on, too; the filling of the shells with the same sum $n + l$ proceeding in order of increasing n. But there are certain deviations from this rule during the filling of the d and f shells.

In the shells with $n = 1, 2, 3$ there are altogether $2 \times 1^2 + 2 \times 2^2 + 2 \times 3^2 = 2 + 8 + 18 = 28$ electrons. There are further eight electrons in the $4s$ and $4p$ states, and another two electrons in the $5s$ state. The $5s$ state is followed by electrons with $n +$

PERIODIC TABLE

Element	$n = 1$	$n = 2$		$n = 3$		Ground State of Atom
	$l = 0$	$l = 0$	$l = 1$	$l = 0$	$l = 1$	
H	1					$^2S^g_{1/2}$
He	2					$^1S^g_0$
Li	2	1				$^2S^g_{1/2}$
Be	2	2				$^1S^g_0$
B	2	2	1			$^2P^u_{1/2}$
C	2	2	2			$^3P^g_0$
N	2	2	3			$^4S^u_{3/2}$
O	2	2	4			$^3P^g_2$
F	2	2	5			$^2P^u_{3/2}$
Ne	2	2	6			$^1S^g_0$
Na	2	2	6	1		$^2S^g_{1/2}$
Mg	2	2	6	2		$^1S^g_0$
Al	2	2	6	2	1	$^2P^u_{1/2}$
Si	2	2	6	2	2	$^3P^g_0$
P	2	2	6	2	3	$^4S^u_{3/2}$
S	2	2	6	2	4	$^3P^g_2$
Cl	2	2	6	2	5	$^2P^u_{3/2}$
Ar	2	2	6	2	6	$^1S^g_0$

$+ l = 6$, where we begin with the least n, that is, with $4d$. There are $2 (4 + 1) = 10$ more of these electrons. The $4d$-electrons are followed by $5p$-electrons, of which there are six, and then by the same rule we get the $6s$ state.

The next value of $n + l = 7$, the least being $n = 4$. Hence, beginning with the 57th place (in actuality, with the 58th place) the $4f$ shell can begin to fill acquiring at once two $4f$-electrons.

This agrees well with the result obtained on the basis of the Thomas-Fermi method for the appearance of electrons with $l = 3$. We saw that electrons with the maximum angular momentum values appear first in the middle of the atom, at the value of the dimensionless variable $x = 0.488$ (this, of course, is an extremely rough estimate).

The same can be said in considering the motion of electrons in a central field decreasing not according to the Coulomb law, but faster, due to screening of the field by other electrons. Specifically, the Thomas-Fermi potential decreases according to the law $Ze/r \, \psi(x)$, approximately as r^{-4}, that is, faster than the centrifugal energy $h^2 l \, (l + 1)/(2mr^2)$. If we add the potential energy of the electron, calculated with due account of screening, and the centrifugal energy, we find that in the d and f shells the minimum of the effective total potential energy lies at the middle of the atom (cf. Sec. 5).

The curve U_M for d- and f-electrons at large r goes higher than for s- and p-electrons, and it turns out that the effective potential well—the minimum on the curve U_M—lies closer to the nucleus than the boundaries of the s- and p-electron shells. Thus, filling of the d and f shells occurs as it were within the atom. But the chemical properties of atoms depend primarily on the outer electrons, the states of which change but slightly in the filling of the f shell. This is how the group of $2 (2 \times 3 + 1) = 14$ chemically similar elements, known as the *rare-earth elements*, is formed.

In the filling of the $3d$ shell, interaction with the outer electrons is stronger; as a consequence, instead of a group of similar elements there appears a series of elements with irregular variations of chemical properties. A "contest" as it were takes place between the $3d$ shell and the outer shell for the energetically most preferable state: for example, V_{23} has three d-electrons and two s-electrons, the next element, Cr_{24}, has five d-electrons and one s-electron, while Mn_{25} also has five d-electrons, but two s-electrons.

Filling of the $5f$ shell takes place, starting with thorium, for a whole group of elements similar to the rare earths. Most of these are artificially produced transuranium elements.

Nuclear Shells. The electron configurations of the noble gases correspond to an especially large binding energy of the electrons in the atom. More work must be done to remove one electron from

the atom of a noble gas than from the atom of any other element. When one more electron is added to the electron shell of the noble-gas type, as in alkali metals, it is very weakly bound.

The noble gases occupy very specific places in the Mendeleyev Periodic Table: 2, 10, 18, 36, etc.; this is explained by the model of filling electron shells, which are in turn formed according to the quantum numbers of individual electrons. But for the concept of quantum numbers of a separate electron to be meaningful we must assume that the electron is subjected to the action of the self-consistent field of all the other electrons.

As was pointed out before, the self-consistent field method as yet remains without any strict substantiation in atomic physics. It is usually said that the Coulomb electrostatic forces act at a distance and therefore each electron is indeed subject to the influence of all the others.

It has been proved experimentally that atomic nuclei are also characterized by especially stable states for each "variety" of specific numbers of nucleons (i.e., neutrons and protons) similar to the ground state of noble gases. These numbers are 2, 8, 20, 50, 82, and 126. But it is well known that, unlike Coulomb forces, nuclear forces act at short range and therefore the argument in favour of the applicability of the self-consistent field concept to the atom is irrelevant with respect to the nucleus.

But since the numbers listed above nevertheless do manifest themselves in very many experimentally observable properties of nuclei, the tongue-in-cheek name of "magic numbers" was suggested to stress their incompatibility with theoretical expectation. Apparently, the self-consistent field concept is nevertheless applicable to the nucleus. Without going into the possible reasons for this, we shall go through the reasoning making it possible to deduce these numbers from a certain simple theoretical model.

It was pointed out in Section 28 that the forces acting between the particles in a nucleus can be described with the help of the effective potential well. Suppose that a certain self-consistent field does nevertheless exist in the nucleus, and it causes each nucleon to move in such a well. The problem of particle motion in a rectangular well was solved in Section 28, and it was shown that it refers not to the one-dimensional case but, in effect, to the three-dimensional case, only as applied to an orbital angular momentum eigenvalue equal to zero.

Such a problem can also be solved for an arbitrary orbital angular momentum of a particle. In order to approximate the solution somewhat closer to reality, we take not a strictly rectangular well, as in Figure 31, but one with slightly rounded edges. Then the energy levels are arranged in groups. Each level in the nucleus is conventionally denoted by the radial, rather than the principal, quantum

number, that is by the number of zeros of the radial function, and by the orbital quantum number. These groups have been found to be: $1s$, $1p$; $1d$, $2s$; $1f$, $2p$; $1g$, $2d$, $3s$; $1h$, $2f$, $3p$; $1i$, $2g$, $3d$, $4s$. (We recall that the letters g, h, i denote the respective angular momenta $l = 4$, 5, 6. There may be 2 $(2l + 1)$ particles in a state with each l.)

Therefore, if we try to correlate the listed groups of separate energy states with the respective shells, we obtain the following numbers of occupied states: 2; $2 + 6 = 8$; $8 + 10 + 2 = 20$; $20 + 14 + 6 = 40$; $40 + 18 + 10 + 2 = 70$; $70 + 22 + 14 + 6 = 112$. They agree with the magic numbers only up to 20.

M. G. Mayer and H. E. Suess explained this discrepancy by the fact that the strong spin-orbit coupling of nucleons was not taken into account in developing the energy levels. The existence of this coupling is well known from nuclear physics, but we shall not go into it. Mayer and Suess postulated that for large l's the state with angular momentum $j = l + 1/2$ differs in energy so greatly from the state with $j = l - 1/2$ that it falls within the states of the preceding group. In that case the states should be grouped as follows (starting with the fourth group): $1f$, $2p$, $1g_{9/2}$; $1g_{7/2}$, $2d$, $3s$, $1h_{11/2}$; $1h_{9/2}$, $2f$, $3p$, $1i_{13/2}$. Since in states with $j = l \pm 1/2$ the spin is rigidly coupled with the orbital angular momentum, they are $(2j + 1)$-fold degenerate. Therefore, after number 20, corresponding to the filling of the first three groups, the next occupied group occurs for $20 + 14 + 6 + 10 = 50$; then for $50 + 8 + 10 + 2 + 12 = 82$; and for $82 + 10 + 14 + 6 + 14 = 126$. This corresponds exactly to the magic numbers.[12]

The suggested explanation leads in turn to certain predictions, namely that close to the magic numbers one should expect nuclei with large angular momenta: $j = 9/2$, $11/2$, . . ., which is in fact the case.

The Ortho- and Para-States of Two Electrons. We shall now show how to develop a spin wave function in accordance with exchange-symmetry requirements. We shall assume that the total wave function separates into a product of the coordinate and spin parts and consider the spin part separately. Since the whole product is antisymmetric, one of its factors must be symmetric and the other antisymmetric. As mentioned before, this simple conclusion refers only to a two-electron system.

Since, in atomic units, the spin of each electron is equal to 1/2, the total spin of a two-electron system may be either zero or unity. These states have special names: the state with unity spin is called the *ortho-state*, the one with zero spin is the *para-state*.

---

[12] Close to the magic numbers is the number 28, which is obtained if the state $f_{7/2}$ is separated from the fourth group.

Neglecting the magnetic interaction between the spins, we can assume that the spin wave function of two electrons can be developed from the products of the spin functions taken separately, in the same way as was done for the coordinate wave function (33.17):

$$\chi(\sigma_1, s_1; \sigma_2, s_2) = \chi(\sigma_1, s_1)\, \chi(\sigma_2, s_2) \qquad (33.41)$$

Now the symmetry requirements must be satisfied. If $\sigma_1 = \sigma_2$, the product (33.41) is intrinsically symmetric. If $\sigma_1 \neq \sigma_2$, then either a symmetric or antisymmetric wave function can be formed, the same as in (33.17). As a result we obtain three symmetric wave functions:

$$\chi(1, s_1)\, \chi(1, s_2)$$

$$\frac{1}{\sqrt{2}}\,[\chi(1, s_1)\, \chi(-1, s_2) + \chi(1, s_2)\, \chi(-1, s_1)]$$

$$\chi(-1, s_1)\, \chi(-1, s_2) \qquad (33.42a)$$

and one antisymmetric:

$$\frac{1}{\sqrt{2}}\,[\chi(1, s_1)\, \chi(-1, s_2) - \chi(1, s_2)\, \chi(-1, s_1)] \qquad (33.42b)$$

The factor $1/\sqrt{2}$ is introduced for normalization.

To the functions (33.42a) there correspond three projections of the total spin: 1, 0, and −1, while only one zero projection corresponds to the function (33.42b). The value of the spin projection, that is, 0, 1, or −1, depends on the choice of the z axis. But the symmetry or antisymmetry of a wave function is an intrinsic property independent of the choice of coordinate axes. Therefore, all three wave functions (33.42a) must be considered as belonging to one and the same value of total spin, equal to 1, but to three different projections; with state (33.42b) belonging to the zero total spin value. The total spin, like the wave function symmetry, does not depend upon the choice of coordinate axes.

The division of spin states into ortho- and para-states holds for all systems comprising two identical particles, and for all spin values. The state with a symmetric spin wave function is the ortho-state, that with the antisymmetrical wave function is the para-state. But at spin values greater than 1/2 there is no one-to-one correspondence between the para- and ortho-states and the total spin. At 1/2 spin, though, the spin function automatically belongs to either the ortho- or the para-state, provided we require that it be the eigenfunction of the total spin operator. In that case it is no longer essential that the particles be identical, provided each possesses 1/2 spin. Thus, the ephemeral electron-positron (positive electron) formation, that is, a system of two different particles, may be in the ortho- or para-state, depending upon the value of the total spin.

The self-consistent field equations involve the expression for the exchange energy (see (33.22a) and (33.22b)) with two signs, depending upon the symmetry of the spatial wave function. But since the symmetry of this function is the reverse of the symmetry of the spin wave function, we can say that the sign of the exchange energy is associated with the state of the particles: "+" in the para-state, and "−" in the ortho-state. The para- and ortho-states, in turn, correspond to the different values of the total spin. Therefore, the sign of the exchange energy may be related to the eigenvalue of the total spin of two electrons.

In Exercise 2, Section 30, it was shown that the eigenvalue of the operator $(\hat{\sigma}_1 \cdot \hat{\sigma}_2)$ is equal to −3 for antiparallel spins and to 1 for parallel spins. Hence, if we introduce the operator $(1 + \hat{\sigma}_1 \hat{\sigma}_2)/2$, its eigenvalues are +1 in the ortho-state and −1 in the para-state. And since the symmetry of the spatial wave function is the opposite of the symmetry of the spin function, we can simply introduce into the Hamiltonian of the system, written in the self-consistent field approximation, the exchange energy multiplied by the quantity $-(1 + \hat{\sigma}_1 \hat{\sigma}_2)/2$. Then the correct sign of the exchange energy will be assured in the equations automatically.

Thus, we have as it were the effective interaction between electron spins, which, as was pointed out, is much larger than the magnetic interaction between them.

The Spin-Orbit Interaction of a Separate Electron. Let us develop an operator describing the interaction of an electron's spin with its orbital angular momentum. Obviously, such an operator can be determined only for a self-consistent field since, being integrals of the motion, the angular momenta of separate electrons cannot be stated otherwise than in the assumption that all the other electrons create a static, central-symmetrical field acting on the given electron.

A strict development of a spin-orbit interaction operator is possible only on the basis of the relativistic electron wave equation (Sec. 37). Here we shall restrict ourselves to a semi-intuitive graphic proof.

Let us pass to a reference frame in which the electron is at rest. In this frame it is subject to the action of both an electric and a magnetic field, while in the system in which the nucleus is at rest there is only an electric field. But we cannot make the transfer directly according to the Lorentz transformations for the field components (15.1a) and (15.1b), because the frame in which the nucleus is at rest is noninertial. We shall therefore transform the reference frame gradually, by means of infinitesimal "steps".

Since the noninertiality is due to the interaction of the electron and the nucleus, we may assume that the interaction itself is effected

in infinitesimal "portions" by the elementary electric charge's variation from $e\lambda$ to e $(\lambda + d\lambda)$. Since the Bohr magneton is itself proportional to the charge, its value here is $\lambda\beta$.

In an increase of the magnetic field by $d\mathbf{H}$ the energy operator of the magnetic field acquires an additional term

$$d\hat{V} = -\frac{\lambda\beta}{2}(\hat{\boldsymbol{\sigma}} \cdot d\mathbf{H})$$

At low relative velocities the Lorentz transformations yield

$$d\mathbf{H} = -\frac{1}{c}\mathbf{v} \times d\mathbf{E} = -\frac{1}{mc}\mathbf{p} \times d\mathbf{E}$$

Since the self-consistent field is a central field, the electric field is equal to

$$\mathbf{E} = -\operatorname{grad}\varphi = -\frac{\mathbf{r}}{r}\frac{d\varphi}{dr}$$

and $d\mathbf{E} = \mathbf{E}\,d\lambda$. Substituting this into the formula for the magnetic energy $d\hat{V}$, we find

$$d\hat{V} = -\frac{1}{2mc}\beta\hat{\boldsymbol{\sigma}}(\hat{\mathbf{p}} \times \hat{\mathbf{r}})\frac{1}{r}\frac{d\varphi}{dr}\lambda\,d\lambda$$

Replacing $\hat{\mathbf{r}} \times \hat{\mathbf{p}}$ by $\hat{\mathbf{M}}$, and the magnetic moment by $eh/(mc)$, we obtain

$$d\hat{V} = \frac{eh}{2m^2c^2}(\hat{\boldsymbol{\sigma}} \cdot \hat{\mathbf{M}})\frac{1}{r}\frac{d\varphi}{dr}\lambda\,d\lambda$$

Integrating with respect to λ from 0 to 1, that is, passing to the total value of the interaction, we find the required *spin-orbit interaction Hamiltonian*

$$\hat{V} = \frac{eh}{4m^2c^2}(\hat{\boldsymbol{\sigma}} \cdot \hat{\mathbf{M}})\frac{1}{r}\frac{d\varphi}{dr} \tag{33.43}$$

Let us evaluate the order of magnitude of the coefficient in Eq. (33.43). For that we note that if we take the Thomas-Fermi potential (33.31), then

$$\frac{d\varphi}{dr} = -\frac{Ze}{r^2}\psi + \frac{Ze}{r}\frac{d\psi}{dr} = -\frac{Ze}{r^2}\left(\psi - x\frac{d\psi}{dx}\right)$$

The factor $\psi - x\,(d\psi/dx)$ is of the order of unity and will in future be discarded. But since the Thomas-Fermi variable, x, is itself of the order of unity, we obtain $r \sim h^2/(Z^{1/3}me^2)$. If we express the orbital angular momentum operator $\hat{\mathbf{M}}$ in terms of the dimensionless operator $\hat{\mathbf{l}}$, that is $\hat{\mathbf{M}} = h\hat{\mathbf{l}}$, and $\hat{\boldsymbol{\sigma}}$ in terms of the double dimensionless spin operator, $2\hat{\mathbf{s}}$, and substitute into Eq. (33.43), the numerical factor of $(\hat{\mathbf{s}} \cdot \hat{\mathbf{l}})$ turns out to have the following order

of magnitude:

$$a \sim \left(\frac{Ze^2}{hc}\right)^2 \times \frac{me^4}{h^2} \qquad (33.44)$$

Here, $e^2/(hc)$ is a dimensionless quantity equal to $1/137$, so that $Ze^2/(hc)$ attains 0.6 at $Z \approx 90$; me^4/h^2 is the atomic unit of energy, equal to 27 eV (1 hartree).

The Interaction of $\hat{L}$ and $\hat{S}$. Thus, the operator of the spin-orbit interaction of a separate electron may be written in the form

$$\hat{V} = a\,(\hat{\mathbf{l}} \cdot \hat{\mathbf{s}}) \qquad (33.45)$$

The total spin-orbit interaction Hamiltonian is obtained by summing $\hat{V}$ for all the electrons in a certain shell with total spin and orbital angular momentum. This is done in the following manner. We assume the total spin **S** of the electrons to be fixed, and we average over the spin involved in each term of the spin angular momentum of the separate electrons. There then remains only the projection of the spin of a separate electron on the total spin:

$$\langle \mathbf{s} \rangle = b\hat{\mathbf{S}}$$

The averaging was done over the spin states for the given total spin considered as an operator. Similarly, the orbital angular momenta are averaged for the given total orbital angular momentum $\hat{\mathbf{L}}$, so that as a result we obtain the spin-orbit interaction Hamiltonian

$$\hat{V}_{so} = A\,(\hat{\mathbf{L}} \cdot \hat{\mathbf{S}}) \qquad (33.46)$$

Let us find the eigenvalues of this operator. For that we write

$$J^2 = (\hat{\mathbf{L}} + \hat{\mathbf{S}})^2 = \hat{L}^2 + \hat{S}^2 + 2\,(\hat{\mathbf{L}} \cdot \hat{\mathbf{S}})$$

But $\hat{L}^2 = L\,(L + 1)$, $\hat{S}^2 = S\,(S + 1)$, and $J^2 = J\,(J + 1)$, whence

$$(\hat{\mathbf{L}} \cdot \hat{\mathbf{S}}) = [J\,(J+1) - L\,(L+1) - S\,(S+1)]/2 \qquad (33.47)$$

Thus, the spin-orbit interaction Hamiltonian is diagonal in states with a specified value of the total angular momentum J. For the given L and S there are altogether $2L + 1$ or $2S + 1$ different states with a given J, which are known together as a *multiplet*. Since L and S are the same for all the multiplet components, we find that the energy of a multiplet component with the given value of J is determined only by the term $J\,(J + 1)$.

Now we are in a position to explain Hund's second rule with respect to the value of J corresponding to the least energy in the multiplet.

Let there be a certain shell $(nl)^p$ in which less than half the places are occupied (p is the number of electrons in the shell, $p < 2l + 1$). Then by Hund's first rule, all the spins line up parallel to each other, and $\hat{\mathbf{s}} = \hat{\mathbf{S}}/p$. The constant a in (33.45) is positive. (This

derives from (33.43), taking into account that $d\varphi/dr < 0$ and that the electron charge is negative.) But then it can be seen that the constant A in Eq. (33.46) is also positive and equal to a/p. Hence, from (33.47), to the least energy of the multiplet there corresponds the least total angular momentum $J = |L - S|$.

If the shell is more than half-filled, it is more convenient to examine the "holes" rather than the electrons. The total spin and orbital angular momenta of a fully occupied shell are zero, so that there is no splitting. Each "hole" signifies the removal of one electron from this occupied state, that is, it has negative energy with respect to the zero level. If there are several "holes", their spins are parallel. Therefore the constant A is now negative: it refers to "holes". To the least energy in the multiplet there corresponds the largest total angular momentum $J = L + S$. Thus, Hund's second rule is validated. When $A > 0$, a multiplet is said to be *normal*, when $A < 0$, it is *inverted*. The multiplet splitting itself is called the *fine structure* of the level with given n and l.

Levels with different L and S are spaced at distances of the order of one or several electron volts (see the example of nitrogen). This order of magnitude is explained, as we have seen, by the electrostatic interaction between electrons, which does not involve c in the denominator. Multiplet splitting is of a magnetic nature and is accordingly smaller. Hence the name "fine structure".

The discourse above refers to light atoms: as is apparent from (33.44), in the case of heavy atoms multiplet splitting is of the same order as the electrostatic interaction between electrons. Therefore in heavy atoms another type of coupling than in light atoms is frequently observed: instead of the orbital and spin angular momenta of all the electrons combining into total L and S, the l's and s's of separate electrons combine to form their total angular momentum, j. After that the total angular momenta of individual electrons combine. This is known as j-j coupling. In nuclei, as we have seen, only j-j coupling occurs.

The Atom in an External Magnetic Field (the Zeeman Effect). In examining the behaviour of a system of charges placed in an external magnetic field, a useful point of departure is the concept of Larmor's magnetic moment precession about a field (see Sec. 17). In such precession only the angular momentum component along the field is conserved, the two perpendicular components averaged over the precession motion being zero.

There is an analogous situation in quantum mechanics, with the only difference that angular momentum projections perpendicular to the field do not exist as physical entities. We thus have a simple correspondence between the integrals of the motion in classical and quantum mechanics. Such a corresponding quantity is the projec-

tion of the angular momentum on the magnetic field; it can be called a *quantum integral of the motion.*

An external magnetic field imposed upon an atom disturbs its state in some specific way. Let us write the perturbing Hamiltonian for the case of normal coupling in the atom. Then the orbital moments of the separate electrons combine into a total orbital angular momentum L, which in turn produces a magnetic moment:

$$\hat{\mu}_{\mathrm{orb}} = \frac{e}{2mc}\hat{\mathbf{L}}$$

The spin magnetic moments also combine into a total spin moment:

$$\hat{\mu}_{\mathrm{sp}} = \frac{e}{mc}\hat{\mathbf{S}}$$

(Here, the 2 does not appear in the denominator owing to the magnetic spin anomaly: see Section 30.)

The total magnetic moment of the atom is

$$\hat{\mu} = \hat{\mu}_{\mathrm{ord}} + \hat{\mu}_{\mathrm{sp}} = \frac{e}{2mc}(\hat{\mathbf{L}} + 2\hat{\mathbf{S}}) \tag{33.48}$$

It follows that magnetic moments combine according to a different law than angular momenta, $\mathbf{J} = \mathbf{L} + \mathbf{S}$. The magnetic moment of an atom with nonzero L and S is not proportional to its angular momentum.

The perturbing Hamiltonian due to the magnetic field is thus equal to

$$\hat{V} = -(\hat{\mu}\cdot\hat{\mathbf{H}})| = \frac{e h \mathbf{H}}{2mc}(\hat{\mathbf{L}} + 2\hat{\mathbf{S}}) = \beta \mathbf{H}(\hat{\mathbf{J}} + \hat{\mathbf{S}}) \tag{33.49}$$

Here, $\beta = eh/(2mc)$ is the Bohr magneton. The "+" sign before it is due to the fact that the electron's charge is negative. We recall that in Eq. (17.33) the magnetic energy was defined precisely as a correction to the Hamiltonian, that is, to the energy expressed in terms of the momenta. That is why in quantum theory it is directly interpreted as an operator.

In the subsequent reasoning it will be more useful to apply the so-called *vector model* of the atom, which makes the conclusions extremely graphic. The idea of this model is that the quantities J, L and S in Eq. (33.47) correspond to the three sides of a triangle. To each multiplet level corresponds its own triangle. In the limiting cases of $J = L + S$ and $J = |L - S|$, the triangle degenerates into a straight line.

Suppose, now, that a constant, homogeneous magnetic field is applied to the atom, and it acts on the atom's magnetic moment according to Eq. (33.49). Two limiting cases are possible in which the vector model offers a very simple picture of the state of the atom due to the effect of the field.

(i) *Weak fields.* Let the external fields be very small in comparison with the effective internal "field" in the atom which causes the whole triangle JLS to rotate about side J. More strictly speaking, the additional energy of the atom due to such a weak field is very small in comparison with the distance between the individual components of the multiplet (both definitions mean basically the same thing).

In Section 32 it was shown that a perturbation may be considered weak if the displacement of levels due to it is small in comparison with the distances between the levels. In the present case this refers to the fine structure levels. Furthermore, the correction to the value of the energy level is obtained by averaging the perturbation energy over the undisturbed state. But we have said that quantum mechanical averaging over a state can be represented as averaging over precessional motion. This is the mode of expression in using the vector model. Of course, the model can offer no more than follows from the general propositions of quantum mechanics. In particular, the precession should not be imagined as a real rotation of the triangle; it is only an indication of how to perform the averaging correctly.

The state of the atom with angular momentum J is $(2J + 1)$-fold degenerate, in accordance with the number of possible J_z projections. Let us now apply the general rule for finding corrections to the energy levels due to the perturbation. From Eq. (32.22), the correction to the level is calculated with the help of matrix elements of the perturbing energy taken between separate undisturbed degenerate states. In the special case, when there are only the matrix elements between the same degenerate states, the energy correction is simply equal to the diagonal matrix element, $V_{n\lambda\,n\lambda}$.

In this case the undisturbed state is degenerate in the angular momentum projection J_z, but the splitting components in a weak field correspond to the same J ($J \equiv n$, $J_z \equiv \lambda$). The first term in the perturbing Hamiltonian (33.49) is J_z itself (**H** is directed along z). Let us show that the second term (S_z) possesses only matrix elements diagonal in J_z, if we take them between states with the same values of the total angular momentum J. And since the magnetic field is by definition weak, it does not alter J.

We write the following operator:

$$\hat{S}_z\,(\hat{J}_x^2 + \hat{J}_y^2 + \hat{J}_z^2) = \hat{J}_z\,(\hat{S}_x\hat{J}_x + \hat{S}_y\hat{J}_y + \hat{S}_z\hat{J}_z)$$
$$+ (\hat{S}_z\hat{J}_x - \hat{J}_z\hat{S}_x)\,\hat{J}_x + (\hat{S}_z\hat{J}_y - \hat{J}_z\hat{S}_y)\,\hat{J}_y$$

Here, the right-hand side of the equation was obtained by means of an identical transformation. The product $\hat{S}_x\hat{J}_x + \hat{S}_y\hat{J}_y + \hat{S}_z\hat{J}_z =$ $= (\hat{\mathbf{S}}\cdot\hat{\mathbf{J}})$ is calculated in the same way as $(\hat{\mathbf{L}}\cdot\hat{\mathbf{S}})$ from (33.47). It is diagonal for the given multiplet component; $\hat{J}_x^2 + \hat{J}_y^2 + \hat{J}_z^2 = \hat{J}^2$

is also diagonal, whence

$$\hat{S}_z = \frac{\hat{J}_z}{2J(J+1)} [J(J+1) - L(L+1) + S(S+1)]$$
$$- \frac{1}{J(J+1)} (\hat{\gamma}_y \hat{J}_x - \hat{\gamma}_x \hat{J}_y)$$

where $\hat{\gamma}$ is determined as follows:

$$\hat{\gamma}_x = \hat{S}_z \hat{J}_y - \hat{J}_z \hat{S}_y = \hat{S}_z \hat{L}_y - \hat{L}_z \hat{S}_y = \hat{L}_y \hat{S}_z - \hat{L}_z \hat{S}_y$$
$$\hat{\gamma}_y = \hat{J}_z \hat{S}_x - \hat{S}_z \hat{J}_x = \hat{L}_z \hat{S}_x - \hat{S}_z \hat{L}_x = \hat{L}_z \hat{S}_x - \hat{L}_x \hat{S}_z$$
$$\hat{\gamma}_z = \hat{L}_x \hat{S}_y - \hat{L}_y \hat{S}_x \tag{33.50}$$

It will be shown in Exercise 4 that $\hat{\gamma}$ has no matrix elements diagonal in J. Consequently, its mean value over the undisturbed state
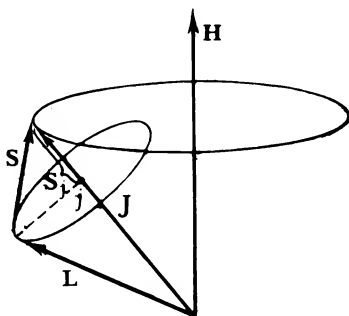
Figure 40

of the system, that is, over one multiplet component corresponding to a definite J, is zero. Thus, $\hat{S}_z$ is proportional to $\hat{J}_z$, at least in the required approximation. From this we immediately obtain the expression for the energy splitting in a magnetic field:

$$E = \beta |\mathbf{H}| J_z \left[1 + \frac{J(J+1) - L(L+1) + S(S+1)}{2J(J+1)} \right] \tag{33.51}$$

The expression in the brackets is known as the *Lande factor*.

Let us now show how Eq. (33.51) is developed on the basis of the vector model. Figure 40 presents a triangle which in the absence of a magnetic field rotates very rapidly around side **J**. In an external magnetic field **H**, which is sufficiently weak, side **J** itself is in a substantially slower precessional rotation about **H**, remaining at a constant angle to it, since we assume J_z to be a constant of the motion; actually this is valid only insofar as J is a constant of the motion, as we have seen from the strict proof.

Owing to the rapid precession of the triangle about **J**, only the projection of **S** on **J**, equal to

$$\mathbf{S}_J = \frac{\mathbf{J}\,(\mathbf{S}\cdot\mathbf{J})}{J^2}$$

remains approximately constant. The term $(\mathbf{S}\cdot\mathbf{J})$ in the numerator of the expression above is found from the equation

$$\hat{L}^2 = (\hat{\mathbf{J}} - \hat{\mathbf{S}})^2 = \hat{J}^2 + \hat{S}^2 - 2\,(\hat{\mathbf{J}}\cdot\hat{\mathbf{S}})$$

whence

$$(\mathbf{J}\cdot\mathbf{S}) = \frac{J\,(J+1) + S\,(S+1) - L\,(L+1)}{2}$$

Substituting S_z for **S** in Eq. (33.49), we again arrive at Eq. (33.51). Thus we obtain a graphic interpretation of the *Lande factor*.

(ii) *Strong fields*. Suppose the magnetic field is very strong, so that the perturbing energy is much larger than the distances between the multiplet components. In terms of the vector model this means that the orbital and spin angular momenta are in much faster precessional motion about **H** than about the third side **J** of the triangle. But as a consequence of magnetic anomaly, the precession of the spin angular momentum is twice as fast, therefore the coupling in the triangle breaks down. This means that in (33.49) it is more convenient to use the first form of notation, in terms of **L** and **S**, rather than **J**, and find the eigenvalues L_z and S_z separately.

The magnetic field breaks the coupling between **L** and **S**, but it is still too weak to lead to transitions between different values of L and S, to which correspond, as we have seen, energy intervals of several electron volts. That is why the eigenvalues of the operator (33.49) are simply obtained if we go through all the different projections L_z and S_z:

$$E = \beta \mid \mathbf{H} \mid (L_z + 2S_z) \qquad (33.52)$$

The two types of multiplet splitting of the multiplet level in a magnetic field manifest themselves in very different ways in atomic spectra, which will be examined in Section 36.

The Atom in a Constant, Homogeneous Electric Field (the Stark Effect). We shall now consider the behaviour of a multiplet level in an external electric field, starting with the case of a weak field, when the shift in the levels due to the field is small in comparison with the natural splitting of the multiplet.

First of all, it should be borne in mind that the angular momentum projection on the electric field is determined only to within sign, because the angular momentum is a pseudovector, while the electric field is a true vector. In a reversal of the signs of all the coordinates,

the signs of the angular momentum projections do not change, where as those of the electric field projections do. But since the choice of a right-handed or left-handed coordinate system is arbitrary, the angular momentum projections on the electric field are physically determined only to within sign.

If J is an integer, there exist $J + 1$ angular momentum projections on the electric field $(0, 1, \ldots, J)$; if J is half-integral, the total number of projections is $J + 1/2$ $(1/2, 3/2, \ldots, J)$. The state with $J = 1/2$ is not split by the electric field at all. Thus, splitting in a magnetic field is more complete.

In a strong electric field the coupling between **L** and **S** breaks down. Then the splitting pattern is as follows. Vector **L** is an integer. It has $L + 1$ projections on the electric field. Since it is coupled more strongly with the field than the spin vector, whose coupling with the electric field is of the same type as that of spin-orbit coupling in a multiplet, the splitting of a level in a strong electric field is primarily determined by the absolute value of the projection on the field L_z. For a given value of the L_z projection the spin projection with respect to L_z already has $2S + 1$ values, since **L** and **S** are both pseudovectors. We obtain $L + 1$ separate groups of $2S + 1$ levels each.

The only exception is the group in which the orbital angular momentum projection on the field, L_z, is zero. In it spin splitting takes place into $S + 1$ or $S + 1/2$ levels, depending on whether S is an integer or half-integer.

The magnitude of the splitting is determined by the relative shift of neighbouring levels. As was shown in Section 31, the level shift equals the mean value of the perturbing energy with respect to the undisturbed motion. From (16.28), we have an expression for the perturbing Hamiltonian in an electric field in the form

$$\hat{V}_{el} = -(\hat{\mathbf{d}} \cdot \mathbf{E}) \qquad (33.53)$$

It can easily be shown that the mean value of this quantity is zero. Indeed, the wave function of an atomic state with a given J is always either odd or even (for the case of the hydrogen atom see below). Therefore the product $\psi_J^* \psi_J$ is necessarily an even function of the coordinates. But then the mean value of the operator (33.53) involves an odd function of the coordinates in the integrand, the dipole moment (cf. (25.19)), so that the whole integral identically becomes zero.

The splitting of levels is obtained only in the second approximation and is therefore quadratic with respect to the external field. But, as will be shown, this refers to a field which does not disrupt the coupling between **L** and **S** (weak fields).

In a hydrogen atom the energy of the electron is determined only by the principal quantum number n and does not depend upon l.

Therefore the state with $E = E_n$ is seen as a superposition of states with different l from 0 to $n - 1$. But for l even, the wave function is even, and for l odd it is odd. Consequently, the function with $E = E_n$ has no definite parity, so that the mean value of the dipole moment does not become zero. That is why in a hydrogen atom the splitting of lines depends upon the electric field linearly.

Note that n and j are involved in the relativistic formula for the energy of a hydrogen atom. For a given j the orbital moment $l = = j \pm 1/2$, and as in the nonrelativistic approximation the state with given n and j has no definite parity, which makes for the linear splitting effect.

Highly excited atomic states always more or less resemble the state of the hydrogen atom, because the nucleus and the atomic core act on an electron far from the nucleus like a point charge. The energy of these states depends upon l according to Eq. (29.51). If the disturbance caused by the field is stronger than the dependence of the energy level upon l, then a linear splitting effect is observed.

A constant electric field not only displaces the energy levels of the atom but quantitatively alters its whole state as well. Let us write the potential energy of an electron in an atom subject to the action of an external electric field $\mathbf{E}$ directed along the z axis:

$$U = U_0 (r) + e \mid \mathbf{E} \mid z \qquad (33.54)$$

For a sufficiently large and negative z, the potential energy far from the atom is less than in the atom, but this domain of values of z is separated from the region of motion of the electron in the atom by a potential barrier. There always exists a nonzero probability of a spontaneous tunneling of the electron through the potential barrier into a free state. Such phenomena were examined in Section 31.

Whatever the state of an atom, in an external electric field it is capable of spontaneous ionization by an electron's penetration of the potential barrier, just as a nucleus disintegrates spontaneously with the emission of an alpha-particle. Of course, if the field is weak the decay probability is negligibly small. The barrier becomes penetrable in a strong field, especially for highly excited atomic states. If the time of spontaneous electron emission in such a state becomes less than the time of quantum emission, the lines in the spectrum corresponding to transitions from this state disappear.

Thus, a perturbation that is weak inside the atom (the atomic unit of field strength $\mid \mathbf{E} \mid = m^2 e^5/h^4 = 5.13 \times 10^9$ V-cm^{-1}, so that an external field is always weak in comparison with the atomic field) nevertheless significantly affects the state, because the conditions at infinity are changed. But if the broadening of atomic levels

due to their finite lifetime with respect to spontaneous ionization is small in comparison with the distance between the levels, the levels may continue to be treated as discrete.

EXERCISES

1. Find the possible states of a system of two *d*-electrons with the same principal quantum numbers.

Solution. Each electron can occur in ten states:

$$A: 2, \frac{1}{2}; \quad B: 1, \frac{1}{2}; \quad C: 0, \frac{1}{2}; \quad D: -1, \frac{1}{2}$$

$$E: -2, \frac{1}{2}; \quad F: 2, -\frac{1}{2}; \quad G: 1, -\frac{1}{2}$$

$$H: 0, -\frac{1}{2}; \quad I: -1, -\frac{1}{2}; \quad J: -2, -\frac{1}{2}$$

The states with positive projections of spin and orbital angular momentum are

AB: 3, 1;	*AC*: 2, 1;	*AD*: 1, 1;	*AE*: 0, 1
AF: 4, 0;	*AG*: 3, 0;	*AH*: 2, 0;	*AI*: 1, 0
AJ: 0, 0;	*BC*: 1, 1;	*BD*: 0, 1;	*BF*: 3, 0
BG: 2, 0;	*BH*: 1, 0;	*BI*: 1, 0;	*CF*: 2, 0
CG: 1, 0;	*CH*: 0, 0;	*DF*: 1, 0;	*DG*: 0, 0;	*EF*: 0, 0

Choosing the states with maximum angular momentum projections, we obtain three resultant states with zero spin:

$$^1S, \ ^1D, \ ^1G, \quad \text{or} \quad ^1S_0^g, \ ^1D_2^g, \ ^1G_2^g$$

and two states with unity spin:

$$^3P, \ ^3F, \quad \text{or} \quad ^3P_2^g, \ ^3P_1^g, \ ^3P_0^g, \quad \text{and} \quad ^3F_4^g, \ ^3F_3^g, \ ^3F_2^g$$

2. Show that in a system of four *p*-electrons with the same principal numbers the states are the same as in a system of two *p*-electrons; in other words, that two electrons have the same states as two "holes".

3. Calculate the total energy of the electrons in an atom according to the Thomas-Fermi method.

Solution. From (28.25), (33.26), and (33.28), the total kinetic energy of all the electrons in an atom is

$$E_{\text{kin}} = \frac{(2m)^{3/2}}{5\pi^2 h^3} \int_0^\infty E_0^{5/2} \times 4\pi r^2 \, dr = \frac{(2m)^{3/2}}{5\pi^2 h^3} \times 4\pi \int_0^\infty (e\varphi)^{5/2} r^2 \, dr$$

because the boundary kinetic energy of the electrons is $e\varphi$. Instead of $e\varphi$ we substitute $Ze^2\psi/r$ and pass to the dimensionless variable x in accordance

with (33.33). Then for E_{kin} we obtain

$$E_{\text{kin}} = \frac{12}{5} \left(\frac{2}{9\pi^2} \right)^{1/3} Z^{7/3} \times \frac{me^4}{h^2} \int_0^\infty \psi^{5/2} \frac{dx}{x^{1/2}}$$

The potential energy falls into two parts: the potential energy of the electrons' interaction with the nucleus

$$E_{\text{pot}}^{(1)} = - \int_0^\infty \frac{Ze^2}{r} \, n \times 4\pi r^2 \, dr$$

and the interaction energy of the electrons among themselves:

$$E_{\text{pot}}^{(2)} = \frac{1}{2} \int_0^\infty \frac{Ze^2}{r} (1 - \psi) \, n \times 4\pi r^2 \, dr$$

The factor 1/2 takes account of the fact that each electron should be counted once. Adding both sides of the potential energy and passing to the dimensionless variable, we obtain

$$E_{\text{pot}} = E_{\text{pot}}^{(1)} + E_{\text{pot}}^{(2)} = -2 \left(\frac{2}{9\pi^2} \right)^{1/3} Z^{7/3}$$

$$\times \frac{me^4}{h^2} \int_0^\infty (\psi^{5/2} + \psi^{3/2}) \frac{dx}{x^{1/2}}$$

The integrals involved in this expression for the energy are easily calculated with the help of Eq. (33.34), namely,

$$\int_0^\infty \psi^{3/2} \frac{dx}{x^{1/2}} = \int_0^\infty \psi'' \, dx = -\psi'(0)$$

because $\psi'(\infty) = 0$. We transform the second integral by parts to get

$$\int_0^\infty \psi^{5/2} \frac{dx}{x^{1/2}} = 2x^{1/2}\psi^{5/2} \Big|_0^\infty - 5 \int_0^\infty x^{1/2}\psi^{3/2}\psi' \, dx$$

$$= -5 \int_0^\infty \psi''\psi' x \, dx = -\frac{5}{2} \int_0^\infty x \frac{d}{dx}(\psi')^2 \, dx$$

$$= -\frac{5}{2} x (\psi')^2 \Big|_0^\infty + \frac{5}{2} \int_0^\infty (\psi')^2 \, dx = \frac{5}{2} \int_0^\infty (\psi')^2 \, dx$$

since the integrated expressions are equal to zero. Further

$$\int_0^\infty (\psi')^2 \, dx = \psi\psi' \Big|_0^\infty - \int_0^\infty \psi\psi'' \, dx = -\psi'(0) - \int_0^\infty \psi^{5/2} \frac{dx}{x^{1/2}}$$

since $\psi(0) = 1$. Hence

$$\int_0^\infty \psi^{5/2}\,\frac{dx}{x^{1/2}} = -\frac{5}{7}\,\psi'(0)$$

Substituting these expressions for the integrals into E_{kin} and E_{pot}, we note that $E_{pot} = -2E_{kin}$, so that the total energy is equal to $-E_{kin}$, in agreement with the result of the exact theory set forth in Exercise 1, Section 32. The quantity $\psi'(0)$ is equal to -1.589. From this we obtain the following formula for the total coupling energy of all the electrons of an atom:

$$E = -0.769\,\frac{me^4}{h^2}\,Z^{7/3} = -20.94Z^{7/3}\ \text{eV}$$

For example, for uranium $E = -8 \times 10^5$ eV, or $-1,6mc^2$.

The dependence $Z^{7/3}$ is easily obtainable without computations in the following way. The Coulomb forces decrease slowly with distance. Therefore all the electrons interact in pairs. There are Z^2 pairs. The mean distance between the electrons decreases as $Z^{-1/3}$ (see (33.33)). This yields $Z^{7/3}$. Note that for nuclei the coupling energy is proportional to the first power of the number of particles (within broad limits). This is an indication of the small distances at which nuclear forces act: each nucleon, that is, proton or neutron, interacts not with all the other nucleons but only with its immediate neighbours.

4. Determine the nonzero matrix elements of $\hat{\gamma}$ defined by Eqs. (33.50) with respect to J and J_z, that is $(\gamma)_{JJ_zJ'J_z'}$.

Solution. We find the computation relations for the components of the total angular momentum and γ. For example

$$\hat{J}_x\hat{\gamma}_y - \hat{\gamma}_y\hat{J}_x = \hat{J}_x(\hat{L}_z\hat{S}_x - \hat{L}_x\hat{S}_z) - (\hat{L}_z\hat{S}_x - \hat{L}_x\hat{S}_z)\,\hat{J}_x$$
$$= (\hat{L}_x + \hat{S}_x)(\hat{L}_z\hat{S}_x - \hat{L}_x\hat{S}_z) - (\hat{L}_z\hat{S}_x - \hat{L}_x\hat{S}_z)(\hat{L}_x + \hat{S}_x)$$
$$= -i\hat{L}_y\hat{S}_x + i\hat{L}_x\hat{S}_y = i\hat{\gamma}_z$$

In the same way we obtain the following commutation relations:

$$\hat{J}_y\hat{\gamma}_x - \hat{\gamma}_x\hat{J}_y = -i\hat{\gamma}_z \qquad \hat{J}_z\hat{\gamma}_x - \hat{\gamma}_x\hat{J}_z = i\hat{\gamma}_y$$
$$\hat{J}_x\hat{\gamma}_z - \hat{\gamma}_z\hat{J}_x = -i\hat{\gamma}_y \qquad \hat{J}_y\hat{\gamma}_z - \hat{\gamma}_z\hat{J}_y = i\hat{\gamma}_x$$
$$\hat{J}_z\hat{\gamma}_y - \hat{\gamma}_y\hat{J}_z = -i\hat{\gamma}_x$$

The like components of vectors $\hat{\mathbf{J}}$ and $\hat{\boldsymbol{\gamma}}$ commute. It can easily be verified that these commutation relations are the same as the commutation relations for the orbital angular momentum components of a particle, $\hat{\mathbf{M}}$, with Cartesian coordinates.

Since $\hat{J}_z$ commutes with $\hat{\gamma}_z$, they are diagonal in the same representation; hence, only the matrix elements $(\gamma_z)_{J_zJ_z'}$ with the same J_z are nonzero.

Further, we take the commutation relations for $\hat{J}_z$ and the components $\hat{\gamma}_x$ and $\hat{\gamma}_y$. We multiply the second by $\pm i$ and add with the first, obtaining

$$\hat{J}_z \,(\hat{\gamma}_x \pm i\hat{\gamma}_y) - (\hat{\gamma}_x \pm i\hat{\gamma}_y)\,\hat{J}_z = \pm \,(\hat{\gamma}_x \pm i\hat{\gamma}_y)$$

We denote $\hat{\gamma}_{\pm} = (\hat{\gamma}_x \pm i\hat{\gamma}_y)$ and take the matrix element of the commutation $\hat{J}_z\gamma_{\pm} - \hat{\gamma}_{\pm}\hat{J}_z = \pm\hat{\gamma}_{\pm}$. Since $\hat{J}_z$ is diagonal, we have

$$J_z\,(\gamma_{\pm})_{J_z J_z'} - (\gamma_{\pm})_{J_z J_z'}J_z' = \pm\,(\gamma_{\pm})_{J_z J_z'}$$

Transferring all the terms to the left-hand side of the equation, we find

$$(J_z - J_z' \mp 1)\,(\gamma_{\pm})_{J_z J_z'} = 0$$

We see from this that the matrix element $(\gamma_{\pm})_{J_z J_z'}$ is not zero only for $J_z' = J_z \pm 1$.

The calculations presented here show how to find nonzero matrix elements. A commutator should be developed between an operator which is regarded as diagonal in the given representation and the operator whose matrix elements are being investigated. Then this matrix element can be taken outside the parentheses. For the matrix element itself not to vanish the expression in parentheses should be zero.

For this we must sometimes perform the commutation twice. Let us examine this case. We develop the commutator

$$\begin{aligned}
[\hat{J}^2\hat{\gamma}_z] &\equiv (\hat{J}_x^2+\hat{J}_y^2+\hat{J}_z^2)\,\hat{\gamma}_z - \hat{\gamma}_z\,(\hat{J}_x^2+\hat{J}_y^2+\hat{J}_z^2)\\
&= (\hat{J}_x^2+\hat{J}_y^2)\,\hat{\gamma}_z - \hat{\gamma}_z\,(\hat{J}_x^2+\hat{J}_y^2)\\
&= \hat{J}_x^2\hat{\gamma}_z - \hat{J}_x\hat{\gamma}_z\hat{J}_x + \hat{J}_x\hat{\gamma}_z\hat{J}_x - \hat{\gamma}_z\hat{J}_x^2\\
&\qquad\qquad + \hat{J}_y^2\hat{\gamma}_z - \hat{J}_y\hat{\gamma}_z\hat{J}_y + \hat{J}_y\hat{\gamma}_z\hat{J}_y - \hat{\gamma}_z\hat{J}_y^2\\
&= \hat{J}_x\,(\hat{J}_x\hat{\gamma}_z - \hat{\gamma}_z\hat{J}_x) + (\hat{J}_x\hat{\gamma}_z - \hat{\gamma}_z\hat{J}_x)\,\hat{J}_x\\
&\qquad\qquad + \hat{J}_y\,(\hat{J}_y\hat{\gamma}_z - \hat{\gamma}_z\hat{J}_y) + (\hat{J}_y\hat{\gamma}_z - \hat{\gamma}_z\hat{J}_y)\,\hat{J}_y\\
&= -i\hat{J}_x\hat{\gamma}_y - i\hat{\gamma}_y\hat{J}_x + i\hat{J}_y\hat{\gamma}_x + i\hat{\gamma}_x\hat{J}_y\\
&= 2i\hat{J}_y\hat{\gamma}_x + \hat{\gamma}_z - 2i\hat{J}_x\hat{\gamma}_y - \hat{\gamma}_z\\
&= 2i\,(\hat{J}_y\hat{\gamma}_x - \hat{J}_x\hat{\gamma}_y)
\end{aligned}$$

By analogy, we write two more commutators:

$$[\hat{J}^2\hat{\gamma}_x] = 2i\,(\hat{J}_z\hat{\gamma}_y - \hat{J}_y\hat{\gamma}_z)$$
$$[\hat{J}^2\hat{\gamma}_y] = 2i\,(\hat{J}_x\hat{\gamma}_z - \hat{J}_z\hat{\gamma}_x)$$

We now find the second commutator:

$$\begin{aligned}
[\hat{J}^2\,[\hat{J}^2\hat{\gamma}_z]] &= 2i\,(\hat{J}_y\,[\hat{J}^2\hat{\gamma}_x] - \hat{J}_x\,[\hat{J}^2\hat{\gamma}_y])\\
&= 4\hat{J}^2\hat{\gamma}_z - 4\hat{J}_y\hat{J}_z\hat{\gamma}_y - 4\hat{J}_x\hat{J}_z\hat{\gamma}_x - 4\hat{J}_z^2\hat{\gamma}_z
\end{aligned}$$

Here, we should take advantage of the fact that all the components of $\hat{\mathbf{J}}$ commute with $\hat{J}^2$. Here we substituted the expressions for the first commutators $[\hat{J}^2\hat{\gamma}_x]$ and $[\hat{J}^2\hat{\gamma}_y]$ into the right-hand side. We make use of the commutation relation for the angular momentum component; then the second commutator reduces to

$$[\hat{J}^2\,[\hat{J}^2\hat{\gamma}_z]] = 4\hat{J}^2\hat{\gamma}_z - 4\,(\hat{J}_z\hat{J}_y + i\hat{J}_x)\,\hat{\gamma}_y - 4\,(\hat{J}_z\hat{J}_x - i\hat{J}_y)\,\hat{\gamma}_x - 4\hat{J}_z^2\hat{\gamma}_z$$

$$= 4\hat{J}^2\gamma_z - 4i\,(\hat{J}_x\hat{\gamma}_y - \hat{J}_y\hat{\gamma}_x) - 4\hat{J}_z\,(\hat{\mathbf{J}}\cdot\hat{\boldsymbol{\gamma}})$$

$$= 4\hat{J}^2\hat{\gamma}_z - 2\,(\hat{J}^2\hat{\gamma}_z - \hat{\gamma}_z\hat{J}^2) - 4\hat{J}_z\,(\hat{\mathbf{J}}\cdot\hat{\boldsymbol{\gamma}})$$

But $(\hat{\mathbf{J}}\cdot\hat{\boldsymbol{\gamma}})$ is identically zero, so that

$$[\hat{J}^2\,[\hat{J}^2\hat{\gamma}_z]] = 2\hat{J}^2\hat{\gamma}_z + 2\hat{\gamma}_z\hat{J}^2$$

Thus we have obtained the relationship between the operators containing only the diagonal operator $\hat{J}^2$ and $\hat{\gamma}_z$:

$$(\hat{J}^2)^2\,\hat{\gamma}_z - 2\hat{J}^2\hat{\gamma}_z\hat{J}^2 + \hat{\gamma}_z\,(\hat{J}^2)^2 - 2\hat{J}^2\hat{\gamma}_z - 2\hat{\gamma}_z\hat{J}^2 = 0$$

From the obtained equation we now form the matrix element with indices J and J'. In this, instead of $\hat{J}^2$ and $\hat{J}'^2$ we must write $J\,(J+1)$ and $J'\,(J'+1)$:

$$J^2\,(J+1)^2\,(\gamma_z)_{JJ'} - 2J\,(J+1)\,(\gamma_z)_{JJ'}\,J'\,(J'+1)$$

$$+ (\gamma_z)_{JJ'}\,J'^2\,(J'+1)^2 - 2J\,(J+1)\,(\gamma_z)_{JJ'} - 2\,(\gamma_z)_{JJ'}\,J'\,(J'+1) = 0$$

We transform the factor of $(\gamma_z)_{JJ'}$ thus:

$$J^2\,(J+1)^2 - 2J\,(J+1)\,J'\,(J'+1)$$

$$+ J'^2\,(J'+1)^2 - 2J\,(J+1) - 2J'\,(J'+1)$$

$$= J^4 + 2J^3 + J^2 - 2JJ'\,(JJ'+J+J'+1) + J'^4 + 2J'^3$$

$$+ J'^2 - 2J^2 - 2J - 2J'^2 - 2J'$$

$$= (J+J')\,[(J-J')^2\,(J+J')$$

$$+ 2\,(J'^2 - JJ' + J'^2) - 2JJ' - (J+J') - 2]$$

$$= (J+J')\,(J+J'+2)\,(J-J'-1)\,(J-J'+1)$$

Thus we arrive at the equation

$$(J+J'+2)\,(J+J')\,(J-J'-1)\,(J-J'+1) = 0$$

The first factor cannot vanish. The second is equal to zero only if $J = J' = 0$. This can never happen in a multiplet. But furthermore, the conditions for $(\gamma_\pm)_{J_z J_z'}$ preclude the possibility of $J = J' = 0$. It follows that the only nonzero matrix element $(\gamma_z)_{JJ'}$ is obtained for $J' = J \pm 1$. But such a matrix element refers to two different components of the multiplet, so that in averaging over one component it makes no contribution to $\langle S_z \rangle$.

We shall make use of the rules for $(\gamma_z)_{JJ_zJ'J_z'}$ in Section 36.

34

DIATOMIC MOLECULES

Homopolar Bonding. Chemical bonding is, in the final analysis, always due to electrostatic interaction between the nuclei and electrons of two atoms. However, for a comprehensive qualitative as well as quantitative explanation of it, quantum mechanics is essential. Bohr's orbits, which were once used to explain the stability of atoms, lack elementary mechanical stability when applied to molecules, that is, they cannot bind atoms together.

There are two different types of chemical bonds: those in which the electric charge passes in part or in whole from one atom to another, and those in which the atoms remain strictly neutral. The former is known as heteropolar, or ionic, bonding, the latter, as *homopolar*, or *covalent*, *bonding*.

In Section 32 it was stated that the electron shell of an atom of fluorine is one electron short of the fully occupied shell of neon. When fluorine combines with hydrogen, the electron of the hydrogen atom transfers to the fluorine atom, which becomes negatively charged and attracts the hydrogen nucleus (proton). The combination of these two atoms is sufficiently stable, and energy is evolved in the process. Quantum mechanics, of course, is required for a comprehensive calculation of such a system, the same as for all atomic calculations, but from the qualitative point of view, at least, it is clear why a negatively charged particle combines with a positively charged one.

When two hydrogen atoms combine into a hydrogen molecule no charge transfer occurs (experimental data reveals that a hydrogen molecule has no dipole moment). Therefore the simple model of two attracting opposite charges is at the very least inadequate to explain the nature of homopolar bonding. The quantum mechanical explanation of the mechanism of such bonding was offered by W. Heitler and F. London in 1926.

In the zero (initial) approximation of the Heitler-London method the atoms are considered to be independent. Each electron is associated with its nucleus. We denote the nuclei by the letters a and b, and the electrons by the subscript 1 and 2. The interaction between the atoms is not taken into account in the initial approximation. The wave functions of the electrons are, respectively, $\psi_a(\mathbf{r_1})$ and $\psi_b(\mathbf{r_2})$. But it is apparent that the state of the system is degenerate, because a system of the same energy results if we take the electron functions $\psi_a(\mathbf{r_2})$ and $\psi_b(\mathbf{r_1})$. The wave function of the zero approximation for a degenerate state is developed according to the general

rules of perturbation theory (see (32.20)-(32.23)) in such a way as to diagonalize the perturbing Hamiltonian.

Like any Hamiltonian of a two-electron system, a Hamiltonian describing atomic interactions is symmetric with respect to an interchange of the two electrons. Hence, if we take

$$\Psi_S = \psi_a\,(\mathbf{r_1})\,\psi_b\,(\mathbf{r_2}) + \psi_a\,(\mathbf{r_2})\,\psi_b\,(\mathbf{r_1}) \tag{34.1}$$

and

$$\Psi_A = \psi_a\,(\mathbf{r_1})\,\psi_b\,(\mathbf{r_2}) - \psi_a\,(\mathbf{r_2})\,\psi_b\,(\mathbf{r_1}) \tag{34.2}$$

as the zero-approximation functions and, as always, denote the perturbing Hamiltonian $\hat{V}_{12}$, then only the matrix elements

$$(V_{12})_{SS} = \int\,\Psi_S^*\hat{V}_{12}\Psi_S\,dx_1\,dx_2 \tag{34.3}$$

$$(V_{12})_{AA} = \int\,\Psi_A^*\hat{V}_{12}\Psi_A\,dx_1\,dx_2 \tag{34.4}$$

do not vanish.

Both wave functions (34.1) and (34.2) must be made to satisfy Pauli's exclusion principle. For that Ψ_S should be multiplied by an antisymmetric spin function, and Ψ_A by a symmetric spin function. Then the Ψ_S state will correspond to total zero spin, and the Ψ_A state to total spin unity (Sec. 33).

Furthermore, both functions Ψ_S and Ψ_A must be normalized, that is, separated into

$$N_S \equiv \left(\int\,|\,\Psi_S\,|^2\,dx_1\,dx_2\right)^{1/2}$$

$$\text{and} \qquad N_A \equiv \left(\int\,|\,\Psi_A\,|^2\,dx_1\,dx_2\right)^{1/2}$$

In future we may write a scalar argument instead of the vector argument $\mathbf{r}$ in the wave functions, since it is assumed that both hydrogen atoms are in the ground state, with spherical symmetry. Accordingly, the arguments involve four distances, r_{a_1}, r_{b_2}, r_{a_2}, and r_{b_1}, while Ψ_S and Ψ_A should be written as

$$\Psi_S = (N_S)^{-1}\,[\psi\,(r_{a_1})\,\psi\,(r_{b_2}) + \psi\,(r_{b_1})\,\psi\,(r_{a_2})] \tag{34.5}$$

$$\Psi_A = (N_A)^{-1}\,[\psi\,(r_{a_1})\,\psi\,(r_{b_2}) - \psi\,(r_{b_1})\,\psi\,(r_{a_2})] \tag{34.6}$$

Here, ψ is one and the same function, that is, the function of the ground state of the hydrogen atom. From (29.23), it is equal to $e^{-\xi}$ where ξ is the distance from the nucleus in atomic units. There is no need to normalize it, since the functions Ψ_A and Ψ_S are subsequently normalized anyway.

We now write the Schrödinger equation for a hydrogen molecule:

$$\left(-\frac{h^2}{2m_p}\nabla_a^2 - \frac{h^2}{2m_p}\nabla_b^2 - \frac{h^2}{2m}\nabla_1^2 - \frac{h^2}{2m}\nabla_2^2 \right.$$

$$\left. -\frac{e^2}{r_{a_1}} - \frac{e^2}{r_{b_2}} - \frac{e^2}{r_{a_2}} - \frac{e^2}{r_{b_1}} + \frac{e^2}{r_{12}} + \frac{e^2}{r_{ab}} \right)\Psi = E\Psi \quad (34.7)$$

The first two terms describe the motion of the nuclei in the molecule. They involve the proton mass in the denominator and are therefore small in comparison with the terms describing the motions of the electrons. Physically, this means that the nuclei are moving much slower than the electron, so that the electron wave function may be determined for a fixed distance between the nuclei; E depends explicitly on the distance between the nuclei, which is involved as a parameter in the left-hand side of (34.7).

We regard $E(r_{ab})$ as the potential energy of the nuclei for the given distance between them. If this function has a minimum denoting the state of stable equilibrium of the nuclei for the given quantum electronic state, the atoms can combine into a molecule. In future we shall not write the terms corresponding to the kinetic energy of the nuclei—they have to be taken into account when vibrational, rotational, or translational motion of a molecule is considered, but the stable equilibrium configuration, which is determined by the electron state, should be defined without account of $-(h^2/2m_p)\nabla_a^2$ and $-h^2/2m_p\nabla_b^2$.

The perturbing Hamiltonian, which was denoted $\hat{V}_{12}$, is transferred into the second line of Eq. (34.7). Thus, in the first approximation we obtain the corrections to the energy of the unperturbed, ground state of the atoms by substituting the wave functions (34.5) and (34.6) into Eqs. (34.3) and (34.4) and calculating the integrals of the known functions. We should not expect great precision in this method, because when the kinetic energy of the nuclei is discarded, Eq. (34.7) does not involve a small parameter; but in the first, semiquantitative, approximation it is nevertheless possible to encompass the main features of the posed problem.

First of all, it should be expected that a lower energy corresponds to the Ψ_S state than to the Ψ_A state. Indeed, Ψ_A vanishes in the plane perpendicular to line r_{ab} at half the distance between the nuclei, that is, in the median plane. Thus Ψ_A has a node. But Ψ_S has no nodes, whence it follows that a smaller energy corresponds to it.

Integrals (34.3) and (34.4) differ in one term:

$$\pm A = \pm N_A^{-1}N_S^{-1}\int \psi(r_{a_1})\psi(r_{b_2})\left(-\frac{e^2}{r_{a_2}} - \frac{e^2}{r_{b_1}} + \frac{e^2}{r_{12}} + \frac{e^2}{r_{ab}} \right)$$

$$\times \psi(r_{a_2})\psi(r_{b_1})\,dx_1\,dx_2 \quad (34.8)$$

which is similar to the exchange integral in the equation for expressing energy by Fock's self-consistent field method (see (33.20)). But in the present case the integral involves known functions, so that it can be calculated and the curve $E(r_{ab})$ can be plotted. The quantity A in (34.8) is also called an *exchange integral*. The corresponding curve indeed has a minimum for a certain value, r_{0ab}, the depth of the potential well corresponding to the binding energy of a hydrogen molecule to the extent that can be expected in the given approximation. If necessary, agreement with experimental data can be enhanced by applying certain computational methods.

The curve with the minimum corresponds only to a symmetric spatial wave function or to an antisymmetric spin function. We obtain, as in the atom, the effective spin interaction energy, which can be represented with the help of the operator $(\hat{\sigma}_1 \cdot \hat{\sigma}_2)$.

A basic feature of valence forces is that they are capable of becoming saturated. A third hydrogen atom does not combine with a diatomic hydrogen molecule. This follows from the Heitler-London theory: the electron spin in the third atom is necessarily parallel to the spin of one of the two atoms in the molecule, so that the extra atom cannot combine with it. The scheme of mutually saturating spins can be applied to a very large class of compounds, especially organic compounds.

However, another approach to the problem of valence is possible, and in some cases absolutely essential. Let us start again with the hydrogen molecule, not with separate atoms this time, but with atoms brought so close together that their nuclei merge. Obviously, in its electrostatic action on the electrons, such a nucleus is equivalent to a helium nucleus, and the electrons of the atoms form an extremely stable shell, as in helium. If the nuclei are not held at one point, the Coulomb repulsive forces push them some distance apart, and the electron shell stretches. At some distance between the nuclei the force appearing in the deformation of the shell balances the repulsion of the nuclei, thereby forming a stable molecular configuration. Here it is also obvious that a third hydrogen atom cannot join it, since there is no place for its electron in a helium shell.

Let us now consider the oxygen molecule. Every oxygen atom has four electrons in the $2p$ shell and spin unity. As is known from experimental data, an oxygen molecule also has unit spin, consequently spin saturation does not occur. This can be explained by examining the common shell of two oxygen atoms combined in a molecule. Assuming that it has six vacancies as in the initial $2p$ shells of the atoms, we see that in combining the extra two electrons pass into a different state. According to Hund's rule their spins are parallel, which explains the value of the spin of the oxygen molecule.

The Electronic States of Diatomic Molecules. Let us consider the electronic terms of a diatomic molecule. The field of a diatomic molecule has axial, but not spherical symmetry, like the field in an atom. Therefore the total orbital angular momentum of the electrons is not an integral of the motion. Only its projection on the line joining the nuclei is conserved. Like the projection of any orbital angular momentum, it takes on integral values: $\Lambda = 0, 1, 2, \ldots$.

The terms are accordingly denoted as Σ, Π, Δ, $\ldots$. Upper-case Greek letters are used so as not to confuse molecular states with atomic states, for which Latin letters are used. The notation takes only the absolute value of the projection into account.

The Hamiltonian of the electrons in a diatomic molecule possesses mirror symmetry with respect to a plane through the nuclei. If we denote a coordinate on an axis perpendicular to this plane ξ, a substitution of $-\xi$ for ξ leaves the Hamiltonian unaffected. It will readily be seen that the sign of the angular momentum projection on the axis is changed at the same time. Indeed, let η be a coordinate lying in a plane perpendicular to the line joining the nuclei. Then the angular momentum projection on the axis joining the nuclei is $M_\xi = \xi p_\eta - \eta p_\xi$. In a reflection in the plane ξ and p_ξ change their signs, while η and p_η do not, so that M_ξ changes its sign too. For that the projection on the axis must not be zero.

Hence, states with $\Lambda \neq 0$, that is Π, Δ, $\ldots$, are 2-fold degenerate: the angular momentum projection on the axis can have two signs for the same energy.

Matters are different with the Σ states. In a mirror reflection in the plane, the function of such a state can only be multiplied by a certain number, C, because there is only one function. In a second reflection it is again multiplied by that number, that is, already by C^2. But at the same time it must revert to its initial value. Consequently $C^2 = 1$, $C = \pm 1$.

In other words, there are two different Σ states. In one the wave function changes its sign in a reflection in a plane through the nuclei, in the other it does not. The first state is termed *positive* and denoted Σ^+, the second is negative and denoted Σ^-.

These are two quite different electronic states of the molecule, and their energies correspond to quite different potential curves $E^+ (r_{ab})$, $E^- (r_{ab})$. The curves are obtained for different initial states of the atoms combining to form a molecule.

The electrons in a molecule may have a total spin S. If there is an odd number of electrons, then S is equal to at least 1/2. If the spin-orbit interaction is not great, the spin orients freely with respect to the orbital angular momentum, and for a given Λ produces a $(2S + 1)$-fold degeneracy (we recall that spin and orbital angular momenta are pseudovectors, so that in stating Λ we obtain $2S + 1$

projection of S). The number $2S + 1$ is written, as for atomic terms, as a superior prefix attached to Λ: $^{2S+1}\Lambda$.

If a molecule consists of two like atoms, the electron wave function possesses an additional symmetry with respect to the median plane between the nuclei perpendicular to the line joining them. A reflection in this plane does not alter the sign of M_ξ, because here for all Λ's there may be two states, even and odd, which in the term notation are written not as for atoms, but as a subscript: Λ_g, Λ_u.

For very many molecules the ground state is $^1\Sigma^+$, and if they are made up of the same atoms, $^1\Sigma_g^+$. The fact that S is equal to zero is explained by the mutual saturation of spins in homopolar molecules. In heteropolar molecules the shell of one of the atoms may be completely filled, as we saw in the case of HF. Here the spin is also equal to zero. Unlike Π, Δ, . . ., the Σ state has no degeneracy, and hence Σ is preferable for the ground state. Finally, unlike odd and negative states, even and positive states need not necessarily have nodal surfaces on the planes of symmetry. Therefore, the ground states are as a rule even and positive.

However, there are exclusions from the rules: the ground state of an oxygen molecule is $^3\Sigma_g^-$. An acceptable explanation of the reason why the spin of the oxygen molecule is unity was cited above. We shall now show how the negative value of the term can be explained. If the spins have equal projections, the spatial states of the electrons must be different. In the initial atoms these were p states, whose wave functions are Y_1^{-1}, Y_1^0, and Y_1^1. But in the present case we must take Y_1^{-1} and Y_1^1 so as to obtain a zero projection of the angular momentum on the symmetry axis for an antisymmetric spatial function.

Since the total spin is unity, the spin function of both electrons is symmetric with respect to their interchange (Sec. 33). Therefore the spatial function must be antisymmetric.

If it is developed from the functions of the electrons in the atoms it has the form

$$Y_1^1(\vartheta_1, \varphi_1)\, Y_1^{-1}(\vartheta_2, \varphi_2) - Y_1^1(\vartheta_2, \varphi_2)\, Y_1^{-1}(\vartheta_1, \varphi_1)$$
$$= \sin\vartheta_1 \sin\vartheta_2 \left(e^{i(\varphi_1-\varphi_2)} - e^{-i(\varphi_1-\varphi_2)}\right) \quad (34.9)$$

In a reflection in a plane through the nuclei the angle $\varphi = \arctan(\xi/\eta)$ changes its sign together with ξ. Hence the function (34.9) also changes its sign, due to the factor in the parentheses. This explains why the ground function is of necessity Σ^-.

Of course, spherical functions are not exact solutions of the electronic states in a molecule, but such a wave-function characteristic as the number or location of nodal surfaces with respect to the symmetry planes does not depend on the fine form of the force field.

Rotation of Nuclei. Let us now consider the motion of a molecule as a whole, restricting ourselves to a singlet state[13] ($S = 0$).

The initial approximation in the theory of molecules consists in considering the motion of electrons when the position of the nuclei is assumed to be fixed. This makes it possible to determine the energy of the electron term as a function of the distance between the nuclei, $E(r_{ab})$. For the sake of brevity we shall write simply r, and replace E by U, since this function takes the place of potential energy in nuclear motion.

In the next approximation the states of motion of the nuclei for the given function $U(r)$ are defined. In other words, their states averaged over the motion of the electrons are considered.

Let us describe the motion in the centre-of-mass frame of the nuclei. As we know from Section 3, a two-body problem reduces to a problem on the motion of one body of mass equal to the reduced mass of the two bodies, m'. Since the potential energy of the nuclei depends only upon the distance between them, their Hamiltonian is conveniently expressed in polar coordinates:

$$\hat{\mathscr{H}}_n = \frac{\hat{p}_r^2}{2m'} + \frac{h^2 \hat{M}^2}{2m'r^2} + U(r) \tag{34.10}$$

In this formula, $\hat{M}^2$ is not an integral of the motion, since the nuclei of a molecule do not really form a closed system: they interact with the electrons. Whatever molecule we have, its total angular momentum $\hat{K}$ is an integral of the motion, as in every closed system. It is compounded of the total electron angular momentum, $\hat{L}$, and the total nuclear angular momentum, $\hat{M}$. Therefore, in Eq. (34.10) we substitute instead of $\hat{M}^2$ the difference between the total and electron angular momenta of the molecule:

$$\hat{M}^2 = (\hat{K} - \hat{L})^2 \tag{34.11}$$

after which we average this operator over the motion of the electrons.

Consider the means of different terms in Eq. (34.11). We have

$$\langle \hat{M}^2 \rangle = \langle \hat{K}^2 \rangle + \langle \hat{L}^2 \rangle - 2 \langle (\hat{K} \cdot \hat{L}) \rangle \tag{34.12}$$

We note that $\hat{K}^2$ is an exact integral of the motion equal to $K(K+1)$, and there is no need to average it; $\langle \hat{L}^2 \rangle$ does not depend on the motion of the nuclei. There remains the third term, which is depend-

---

[13] Atomic and molecular multiplets are designated as singlets, doublets, triplets, quadruplets (or quarters), and further as in music: quintets, sextets, septets, etc.

ent upon $\hat{\mathbf{K}}$ and $\hat{\mathbf{L}}$. Only the projection of $\hat{\mathbf{L}}$ on the molecular axis ζ equal to Λ is diagonal.

The operators of the projections on the other axes do not possess diagonal elements (see (30.18)). Consequently only the mean value of the projection on the ζ axis, that is, the integral of the motion Λ, is other than zero. It remains to find the mean value $\langle (\hat{\mathbf{K}} \cdot \hat{\mathbf{L}}) \rangle =$ $= (\mathbf{K} \cdot \langle \hat{\mathbf{L}} \rangle)$.

The orbital angular momentum operator is defined as

$$\hat{\mathbf{M}} = \hat{\mathbf{r}} \times \hat{\mathbf{p}}$$

where $\hat{\mathbf{r}}$ is the separation of the nuclei and is by definition directed along the molecular axis, and $\hat{\mathbf{p}}$ is the momentum of the relative motion of the nuclei. Thus

$$(\hat{\mathbf{r}} \cdot \hat{\mathbf{M}}) = \hat{\mathbf{r}} (\hat{\mathbf{r}} \times \hat{\mathbf{p}}) = (\hat{\mathbf{r}} \times \hat{\mathbf{r}}) \hat{\mathbf{p}} = 0$$

identically, or

$$\hat{\mathbf{r}} (\hat{\mathbf{K}} - \hat{\mathbf{L}}) = 0 \qquad \text{and} \qquad (\hat{\mathbf{r}} \cdot \hat{\mathbf{K}}) = (\hat{\mathbf{r}} \cdot \hat{\mathbf{L}}) \qquad (34.13)$$

so that the total angular momentum projection is equal to the electron's angular momentum projection on the same axis, that is, it is equal to Λ. The maximum projection of $\mathbf{K}$ on an arbitrary fixed axis in space will thus take on values starting with Λ, that is Λ, $\Lambda + 1$, $\Lambda + 2$, ... The greatest projection of $\mathbf{K}$ on an arbitrary axis cannot be smaller than Λ.

Thus, the mean value of the scalar product

$$\langle (\hat{\mathbf{K}} \cdot \hat{\mathbf{L}}) \rangle = (\mathbf{K} \cdot \langle \hat{\mathbf{L}} \rangle) = (\mathbf{K} \cdot \mathbf{n}) \Lambda = \Lambda^2 \qquad (34.14)$$

where $\mathbf{n} = \mathbf{r}/r$. Like $\langle \hat{L}^2 \rangle$ it does not depend upon the motion of the nuclei.

From this we find the Hamiltonian of the nuclei, averaged over the motion of the electrons:

$$\langle \hat{\mathcal{H}} \rangle_{\text{el}} = \frac{p_r^2}{2m'} + U(r) + \frac{h^2 (\langle \hat{L}^2 \rangle - 2\Lambda^2)}{2m'r^2} + \frac{h^2 K(K+1)}{2m'r^2} \qquad (34.15)$$

Here it is convenient to include a term depending upon $\langle \hat{L}^2 \rangle$ and Λ^2 in the potential energy by analogy with the way the centrifugal energy is included in U_M of the motion in a central field.

Usually, the equilibrium distance between the nuclei, denoted r_e, is substituted into the denominator of the last term. Then $m'r_e^2$ represents the molecule's moment of inertia about its centre of mass. It is apparent that in a system of two mass points only two, and equal, components of the inertia tensor with respect to the principal axes, drawn through the centre of inertia, are other than zero. The moment of inertia about the symmetry axis is zero.

Energy Levels of a Molecule. Close to the potential energy minimum, a system is capable of small vibrations. The vibration frequency in a diatomic molecule is conventionally denoted ω_{vib}, and the corresponding *vibrational quantum number*, v. The energy of a diatomic molecule may be represented as a sum of three terms:

$$E = E_{\text{el}} + h\omega_{\text{vib}}\left(v + \frac{1}{2}\right) + \frac{h^2 K(K+1)}{2m'r_e^2} \tag{34.16}$$

Here, E_{el} is the energy corresponding to the minimum of the potential curve (including the centrifugal term), and the second term is the *vibrational energy* (cf. (28.51)). (Further on it will be explained that Eq. (34.16) refers only to singlet molecular states.)

Let us compare the order of magnitude of the three terms in (34.16). The first is independent of the mass of the nuclei and is determined only by the electron motion. The distance between the electron energy states is of the order of one or several electron volts.

The frequency of small vibrations in the second term involves the square root of the mass of the oscillating particles in the denominator (cf. (7.12), where α is proportional to the mass). Thus, the electron energy of the molecule is several tens or hundreds of times greater than its vibrational energy. For the hydrogen molecule, which has the smallest reduced mass, the vibrational quantum $h\omega_{\text{vib}}$ is about 0.5 eV; for other diatomic molecules it is $h\omega_{\text{vib}} \sim \sim 0.1\text{-}0.2$ eV.

The third term in (34.16), which is due to the rotation of the molecule as a whole, involves the mass of the nuclei in the denominator. The distances between the energy levels are in this case correspondingly smaller and comparable with the distances between the fine structure levels due to the magnetic interaction of the electrons. Under these conditions magnetic interaction cannot be considered small in comparison with the rotational energy of the molecule, and it requires a special consideration for nonsinglet states. We shall not take up this question, however.

Let us now consider the parity of the states of the molecule as a whole with respect to an inversion of the coordinates of all its particles, that is, the electrons and the nuclei. Inversion corresponds to a transition from a right-handed coordinate system to a left-handed one. In the process, the wave function may either remain unchanged or change sign.

The behaviour of $\Sigma^{\pm}$ in this is connected with the rotational quantum number K. Note, firstly, that the inversion is performed simultaneously in the fixed and intrinsic frames of reference. In the intrinsic frame, the Σ^+ terms remain unchanged, while the Σ^- terms change their sign. Furthermore, the rotation of a molecule as a whole is determined by the wave function, the parity of which depends upon

the parity of K (see Sec. 29). Therefore, the parity of the Σ^+ term is $(-1)^K$, and that of the Σ^- term is $(-1)^{K+1}$.

The Effect of Nuclear Spin. The rotational states of homonuclear molecules depend on the spins of the nuclei. Nuclear spins make for an additional symmetry of the Hamiltonian associated with interchange of the nuclei. Depending upon the spin of each nucleus, their wave function should either reverse its sign in a complete interchange of all the variables (spatial and spin), or remain unchanged. The first corresponds to nuclei with half-integral spin, subject to Pauli's exclusion, the second corresponds to nuclei with integral or zero spin. In this case it is immaterial that nuclei are not elementary particles. Insofar as they are moving separately, the applicability of Pauli's exclusion is determined only by their total spin.

The parity of the rotational wave function of a molecule in the Σ state with respect to a reflection in the median plane between the nuclei depends only upon K. Indeed, such a substitution means that in place of the polar angle ϑ we take $\pi - \vartheta$, so that $\cos \vartheta$ is replaced by $-\cos \vartheta$. The function Y_K^0 is the Kth Legendre polynomial, whose parity is equal to $(-1)^K$. The projection of K on the symmetry axis in the Σ state is zero $(K_s = \Lambda)$, and to this state corresponds the wave function $Y_K^0 = P_K (\cos \vartheta)$. Thus, for even K the spatial wave function of the nuclei is even, and for odd K it is odd. The parity of the spin wave function of identical nuclei in the Σ state of a molecule is opposite the parity of K if the nuclear spin is half-integral, and coincides with the parity of K if the nuclear spin is integral.

Let us first consider the case when the spin of each nucleus is equal to 1/2, as in hydrogen. Then Pauli's exclusion principle holds. A symmetric spin function corresponds to spin 1 (orthohydrogen), an antisymmetrical spin function corresponds to spin 0 (parahydrogen). But spin 1 has three projections, 1, 0, -1, so that the molecules of orthohydrogen are in a 3-fold degenerate state with respect to the spin, while the molecules of parahydrogen are not in a degenerate state.

In equations of the type (28.25) for the number of states of a molecule, the molecules of orthohydrogen must be multiplied by the factor 3, and molecules of parahydrogen by the factor 1. Furthermore, both are multiplied by the factor $2K + 1$, according to the number of projections of K.

The situation changes if we take deuterium instead of hydrogen. The nuclei of deuterium have unity spin, and they are not subject to Pauli's exclusion principle. Their wave function is symmetric with respect to an interchange of the spatial and the spin variables. The parity of the spatial function is determined by the parity of K, while the parity of the spin function is determined as follows: the

states with total spin 0 and 2 are even, the state with unity total spin is odd (see the Exercise). These states are also known as ortho- and para-states, respectively. But unlike hydrogen, deuterium occurs in the ortho-state for even K, and in para-states for odd K.

Homonuclear molecules which do not possess spin may be only in even rotational states, because for them the total parity of the wave function with respect to the interchange of variables is determined solely by the parity of K. We repeat that the above reasoning refers only to the Σ states of nuclei.

The magnetic energy of the interaction of nuclear spins with their rotation is very small: it involves c^2 and the mass of the heavy particle in the denominator. But thanks to the symmetry requirements for the wave function, the spin state of the nucleus strongly affects the motion of the molecule.

The formula for the rotational levels of diatomic molecules according to the energy dependence upon K may be applied to non-spherical nuclei as well. Such nuclei are due to the following circumstance. We have seen that nucleons with large angular momenta, whose wave function is very far from being spherically symmetrical, also take part in the filling of nuclear shells. If such a separate nucleon is moving in the field of a symmetric core, owing to the asymmetry of its own state it deforms the core, which elongates and acquires axial symmetry instead of spherical. The rotational levels of the nuclei correspond to low excitation energies (on the nuclear scale) and can be distinguished because they obey the interval rule

$$E_K - E_{K-1} = \frac{h^2}{I} K \tag{34.17}$$

where I is the moment of inertia of the nucleus.

EXERCISE

Develop the wave functions of the ortho- and para-states of a deuterium molecule in the $^1\Sigma_g^+$ electronic state.

Answer.

(i) Ortho-states. Spin projection 0:

$$\Psi(1, s_1)\Psi(-1, s_2) + \Psi(1, s_2)\Psi(-1, s_1), \qquad \Psi(0, s_1)\Psi(0, s_2)$$

Spin projection ± 1:

$$\Psi(1, s_1)\Psi(0, s_2) + \Psi(1, s_2)\Psi(0, s_1)$$
$$\Psi(-1, s_1)\Psi(0, s_2) + \Psi(-1, s_2)\Psi(0, s_1)$$

Spin projection ± 2:

$$\Psi(1, s_1)\Psi(1, s_2), \qquad \Psi(-1, s_1)\Psi(-1, s_2)$$

(ii) Para-states. Spin projection 0:

$$\Psi(1, s_1)\, \Psi(-1, s_2) - \Psi(1, s_2)\, \Psi(-1, s_1)$$

Spin projection ± 1:

$$\Psi(1, s_1)\, \Psi(0, s_2) - \Psi(1, s_2)\, \Psi(0, s_1)$$
$$\Psi(-1, s_1)\, \Psi(0, s_2) - \Psi(-1, s_2)\, \Psi(0, s_1)$$

35

THE QUANTUM THEORY OF SCATTERING

The concept of the scattering cross section of particles, which was defined in Section 6 in terms of classical mechanics, is directly extended to quantum mechanics. Indeed, the differential *scattering cross section* of the particles inside a given solid angle is the ratio of the number of scattered particles in this angle to the flux density of the incident particles. Since flux and flux density can be defined quantum mechanically, the effective cross section has the same sense in quantum theory as it has in classical theory.

We shall first examine an approximate method of determining the scattering cross section of particles and then go over to more exact methods.

The Born Approximation. Let us determine the conditions in which a scattering field can be regarded as a weak perturbation acting on a particle.

Let the particle's energy at a sufficiently large distance from the scatterer be E, and let the potential energy be of the order of magnitude of U. It can be considered, for example, to be the depth of the potential well representing the scatterer. We shall first consider the case of $E \gg U$. Then the change in the wave vector of the particle in the field is of the order

$$\frac{[2m\,(E-U)]^{1/2}}{h} - \frac{(2mE)^{1/2}}{h} \sim \left(\frac{m}{2E}\right)^{1/2} \frac{U}{h}$$

If the dimensions of the region in which the field acts (the "width of the potential well") are of the order of a, then the total phase change of the wave function in the scattering field is estimated as

$$\left(\frac{m}{2E}\right)^{1/2} \frac{Ua}{h} = \frac{Ua}{hv}$$

where v is the velocity of the particle at a large distance from the scatterer. The obtained ratio must be substantially smaller than unity to consider the disturbance produced by the field weak.

In the opposite case (when $|U| \gg E$) the wave number "inside the well" (more precisely, "above the well", because $E > 0$, that is, the motion of the particle is infinite) is equal to $(2m |U|/h)^{1/2}$.

The criterion of the smallness of the phase change is $a (2m |U|/h)^{1/2} \ll 1$ (this relation does not include the energy of the particle at all). Comparing this with the result of Section 28, which refers to a well of finite depth, we see that the obtained condition almost coincides with the condition for the absence of bound states in the well, but it involves a strong inequality rather than a simple one. A well can be regarded as a weak perturbation of the initial motion of the particle only when the product $a \sqrt{|U|}$ is many times smaller than the quantity at which the appearance of a bound level is possible.

If the scattering is done not by a potential well, but by a potential hump, that is, the particles are subject to repulsive forces, the criterion of the smallness of the perturbation is the same, but it bears no relation to bound states.

When the necessary criteria of the smallness of the perturbation are satisfied, we may apply the general methods of Section 32, namely Eq. (32.42), to the problem of particle scattering. We shall assume the scattering to be elastic, that is, the state of the scatterer does not change in the scattering. Then the energy of a scattered particle prior to the scattering is the same as after the scattering, only the direction of its linear momentum changes.

Assuming the perturbation to be weak, we may take the wave functions of the incident and the scattered particle in the form of plane waves. This corresponds to the *Born approximation*. Let the momentum of a particle prior to the impact be **p**, and after the impact, $-\mathbf{p}'$, where, as just pointed out, $p = p'$. Both momenta are determined in the centre-of-mass frame of reference of the colliding particles, and the two-body problem is, as always, reduced to the one-body problem.

Equation (32.42) involves the function $g(E')$, that is, the number of states per unit energy interval. To make use of Eq. (28.25) for $g(E')$, we must normalize the wave functions of the free particles to the same volume V that is involved in $g(E')$. It is, of course, eliminated from the final result. We could ultimately achieve the same result by normalizing the wave functions to the δ function, that is, to $\delta (\mathbf{p} - \mathbf{p}')$, but we would have to alter Eq. (28.25).

Thus, the wave functions of the initial and final states are:

$$\psi (\mathbf{p}) = \frac{1}{V^{1/2}} e^{i\mathbf{p}\mathbf{r}/h}, \qquad \psi (\mathbf{p}') = \frac{1}{V^{1/2}} e^{i\mathbf{p}'\mathbf{r}/h} \tag{35.1}$$

The function $\psi(\mathbf{p})$ corresponds to a flux density $\mathbf{v}/V$, which can be seen directly from (23.19):

$$\mathbf{j} = \frac{h}{2imV}\left(e^{-i\mathbf{p}\mathbf{r}/h} \quad \text{grad } e^{i\mathbf{p}\mathbf{r}/h} - e^{i\mathbf{p}\mathbf{r}/h} \quad \text{grad } e^{-i\mathbf{p}\mathbf{r}/h}\right)$$

$$= \frac{\mathbf{p}}{mV} = \frac{\mathbf{v}}{V} \tag{35.2}$$

From (35.1), the matrix element for the transition is

$$U_{\mathbf{p}'\mathbf{p}} = \frac{1}{V}\int e^{i(\mathbf{p}-\mathbf{p}')\mathbf{r}/h}U(\mathbf{r})\,dV \tag{35.3}$$

To find the transition probability, this matrix element must be substituted into (32.42) and multiplied by $2\pi/h$ and the number of final states $g\,(E' = E)$. Since we are concerned with the probability of a particle occurring in a solid-angle element $d\Omega$, Eq. (28.25) should be additionally multiplied by $d\Omega/(4\pi)$. Therefore

$$dg\,(E' = E) = \frac{dN\,(E)}{dE}\times\frac{d\Omega}{4\pi} = \frac{Vm^{3/2}E^{1/2}}{2^{5/2}\pi^3 h^3}\,d\Omega \tag{35.4}$$

The differential scattering cross section is equal to the scattering probability in unit time inside the solid-angle element, divided by the flux density of the incident particles, $\mathbf{v}/V \equiv (2E/m)^{1/2}/V$. Therefore

$$d\sigma = \left|\int e^{i(\mathbf{p}-\mathbf{p}')\mathbf{r}/h}U(\mathbf{r})\,dV\right|^2 \frac{m^2}{4\pi^2 h^4}\,d\Omega \tag{35.5}$$

The matrix element appearing here now refers to the normalization of the wave functions to the unit volume, $V = 1$. Introducing the wave vectors $\mathbf{k} \equiv \mathbf{p}/h$, $\mathbf{k}' \equiv \mathbf{p}'/h$, we write

$$U_{\mathbf{k}'\mathbf{k}} = \int e^{i(\mathbf{k}-\mathbf{k}')\mathbf{r}}U(\mathbf{r})\,dV \tag{35.6}$$

$$d\sigma = \frac{m^2}{4\pi^2 h^4}\,|U_{\mathbf{k}'\mathbf{k}}|^2\,d\Omega \tag{35.7}$$

The expression (35.5) is simplified when the scattering field is a central field, that is, it depends only upon r. Then integration can be carried out over the angles in $U_{\mathbf{k}'\mathbf{k}}$, leaving only integration with respect to the radius. In defining the polar angle we choose the direction of $\mathbf{k} - \mathbf{k}'$ as the polar axis. After that the matrix element reduces to the form

$$U_{\mathbf{k}'\mathbf{k}} = \int e^{i(\mathbf{k}-\mathbf{k}')\mathbf{r}}U(r)\,dV$$

$$= 2\pi\int_0^\infty r^2\,dr\,U(r)\int_0^\pi e^{i|\mathbf{k}-\mathbf{k}'|r\cos\vartheta}\sin\vartheta\,d\vartheta \tag{35.8}$$

Since $\sin \vartheta \, d\vartheta = -d \cos \vartheta$, we can integrate with respect to ϑ immediately:

$$U_{\mathbf{k'k}} = 2\pi \int_0^\infty r^2 \, dr \, U(r) \frac{e^{i|\mathbf{k}-\mathbf{k'}|r} - e^{-i|\mathbf{k}-\mathbf{k'}|r}}{i|\mathbf{k}-\mathbf{k'}|r}$$

$$= \frac{4\pi}{|\mathbf{k}-\mathbf{k'}|} \int_0^\infty rU(r) \sin(|\mathbf{k}-\mathbf{k'}|r) \, dr \tag{35.9}$$

As we have already said, $k = k'$. Therefore, the vector difference is easily expressed in terms of the deflection angle θ for the particle:

$$|\mathbf{k}-\mathbf{k'}|^2 = 2k^2 - 2(\mathbf{k} \cdot \mathbf{k'}) = 2k^2(1-\cos\theta)$$

$$= 4k^2 \sin^2 \frac{\theta}{2} \tag{35.10}$$

This can also be seen from a geometrical construction.
Thus

$$U_{\mathbf{k'k}} = \frac{2\pi}{k \sin(\theta/2)} \int_0^\infty rU(r) \sin\left(2kr \sin\frac{\theta}{2}\right) dr \tag{35.11}$$

Take, for example, the Thomas-Fermi potential (33.31). Then

$$U = -\frac{Ze^2}{r} \psi(x)$$

Substituting into (35.11) yields

$$U_{\mathbf{k'k}} = -\frac{2\pi Ze^2}{k \sin(\theta/2)} \int_0^\infty \psi(x) \sin\left(2kr \sin\frac{\theta}{2}\right) dr \tag{35.12}$$

This integral can be determined numerically for different values of the dimensionless ratio

$$\frac{(3\pi)^{2/3}}{2^{7/3}Z^{1/3}} \, k \, \frac{h^2}{me^2}$$

whence, after substitution into the general equation (35.7), we obtain the differential scattering cross section of a fast charged particle in an atom.

Let us consider the limiting case of small scattering angles, for which $\sin(\theta/2)$ is replaced by $\theta/2$. Then (35.12) is reduced to the form

$$U_{\mathbf{k'k}} = -\pi \left(\frac{3\pi Z}{4}\right)^{4/3} \frac{h^4}{m^2 e^2} \int_0^\infty x\psi(x) \, dx \tag{35.13}$$

For $x \to \infty$, the function $\psi(x)$ decreases approximately as x^{-3}, so that the integral involved in (35.13) has a finite value. This means

that at small angles the differential scattering cross section referred to a solid-angle unit, that is $d\sigma/d\Omega$, tends to a finite limit. As can be seen from (35.11), in the most general case for this the integral

$$\int_0^\infty r^2 U(r)\,dr < \infty \qquad (35.14)$$

must converge. Then $d\sigma/d\Omega$ tends to a finite limit at infinitesimal scattering angles.

Thus, the dependence of the scattering cross section on the distance to the scatterer in quantum mechanics is different than in classical mechanics. In Section 6 it was shown that if a force does not identically become zero at infinite distance from the scatterer, and does not simply tend to zero, then every particle, however far from the scatterer it may pass, is slightly deflected. For that reason the classical expression for the differential scattering cross section for a small angle tends to infinity if the force of interaction of the particle with the scatterer does not vanish at some finite distance r_0 from the scatterer.

This difference between the classical and the quantum theories is explained on the basis of the uncertainty principle. The uncertainty in the momentum of a particle passing at a distance r from a scatterer is $\Delta p \sim 2\pi\hbar/r$. If the interaction force decreases fast enough with distance, the change in momentum for a large impact parameter r may prove smaller than Δp. This condition is defined by the integral (35.14): when it converges, the uncertainty in the momentum, Δp, associated with wave diffraction is greater than the deflection as a result of the interaction of the particle with the scatterer. But owing to diffraction the integrals take on finite values.

Suppose that the second condition for the applicability of the Born approximation for the case of slow particles is satisfied. Then $kr \ll 1$, since at small energies the wave number is also small. In the expression (35.11) we replace $\sin[2kr \sin(\theta/2)]$ by $2kr \sin(\theta/2)$. It turns out that if the condition (35.14) is satisfied, then the scattering angle is eliminated altogether from the expression for the scattering cross section. The scattering becomes isotropic. The same number of particles passes through each solid-angle element.

Note that condition (35.14) is sufficient, though not necessary, for the total cross section to be finite, that is $\sigma = \int d\sigma < \infty$. Even when the differential scattering cross section for small angles becomes infinite, but not according to a very strong law, the total cross section may remain finite.

Let us now consider the case of particle scattering in a purely Coulomb field. Then $U = \pm Ze^2/r$. The matrix element involved in the general expression (35.5) was calculated for a Coulomb field

in Section 26 (see (26.37)), but we shall find it directly from (35.9), defining the integral $\int\limits_0^\infty \sin x\, dx$ thus

$$\lim_{\alpha\to 0} \int\limits_0^\infty e^{-\alpha x} \sin x\, dx = \lim_{\alpha\to 0} \frac{1}{\alpha^2+1} = 1$$

Then

$$\int\limits_0^\infty \sin ax\, dx = \frac{1}{a}$$

and finally

$$U_{\mathbf{k'k}} = \pm \frac{2\pi Ze^2}{k\sin(\theta/2)} \int\limits_0^\infty \sin\left(2kr\sin\frac{\theta}{2}\right) dr$$

$$= \pm \frac{\pi Ze^2}{k^2 \sin^2(\theta/2)} \tag{35.15}$$

Substituting this into (35.7), we obtain the final expression for the differential cross section:

$$d\sigma = \frac{Z^2 e^4\, d\Omega}{4m^2 v^4 \sin^4(\theta/2)} \tag{35.16}$$

where we have taken advantage of the fact that $p = \hbar k = mv$. This result curiously agrees with the precise classical *Rutherford formula* (6.21).

It turns out that Eq. (35.16) is also obtained from a precise solution of the wave equation for the case of a Coulomb field in the scattering problem. Thus, the Rutherford formula is extended to quantum mechanics unchanged.

The Born approximation in the theory of scattering by a Coulomb field can be regarded as the first nonvanishing term in a series expansion of the exact formula in powers of Ze^2. But since the exact formula does not involve powers higher than $(Ze^2)^2$, the result of the Born approximation coincided with the precise result.

We shall now estimate the limits of applicability of the method under consideration for the Coulomb field. To do this, we make use of the first criterion, referring to large velocities. Since the product Ua in this case is equal to Ze^2, we arrive at the following condition:

$$\left(\frac{m}{2E}\right)^{1/2} \frac{Ua}{\hbar} = \frac{Ze^2}{\hbar v} \ll 1 \tag{35.17}$$

The quantity $e^2/(\hbar c) = 1/137$. Therefore, we write (35.17) thus:

$$\frac{Z}{137}\frac{c}{v} \ll 1 \tag{35.18}$$

But $Z \sim 90$ for heavy elements, so that (35.18) is not satisfied in general. Of course, the Rutherford formula is applicable to non-relativistic particles in this case too, because it is exact.

Formula (35.16) is applicable also to the scattering of relativistic particles at small angles, provided m is replaced by $m\,(1 - v^2/c^2)^{-1/2}$. The condition (35.18) here is not necessary, since small scattering angles correspond to large impact parameters, for which the Coulomb field represents a small disturbance.

The General Theory of Particle Scattering in a Central Field. We shall now consider the particle scattering problem in its exact form. For this we must solve the Schrödinger equation for the motion of a particle in a central field. But unlike Section 29, where we determined the eigenfunctions of the energy operator, here we must look for solutions corresponding to entirely different boundary conditions. Namely, at infinity the wave function must comprise a function corresponding to an incident plane wave, which describes the incident particles, and a function corresponding to a spherical outgoing wave, which describes the scattered particles. Both functions are complex, and we must make use of expansion methods somewhat differing from the general methods of expanding wave functions in operator eigenfunctions (Sec. 25).

Every wave function of a particle in a central field satisfies the Schrödinger equation

$$-\frac{h^2}{2m}\frac{d^2\Phi}{dr^2} + \frac{\hat{M}^2\Phi}{2mr^2} + U\,(r)\,\Phi = E\Phi \qquad (35.19)$$

Here, Φ denotes the wave function divided by r (cf. (29.14)); the energy E is positive, since the motion is infinite.

We represent the solution Φ in the form of an expansion in eigenfunctions of the angular momentum square. We select for the polar axis the direction of the linear momentum of the incident particles. We also make use of the fact that the scattering is symmetric with respect to the azimuth. If the particles possess no spin, or if their spin state is a mixture of states with spins of opposite sense in equal proportions (see Sec. 27), then there always exists azimuthal scattering symmetry, and such symmetry will exist as long as spin-orbit coupling is of no consequence.

The eigenfunctions of angular momentum with azimuth symmetry are $Y_l^0 = P_l\,(\cos\theta)$ (see Sec. 29). Therefore the general solution to Eq. (35.19) of the required form is

$$\psi = \frac{\Phi}{r} = \frac{1}{r}\sum_l R_l\,(r)\,P_l\,(\cos\theta) \qquad (35.20)$$

where $R_l(r)$ satisfies the equation

$$-\frac{h^2}{2m}\frac{d^2R_l}{dr^2}+\frac{h^2l(l+1)R_l}{2mr^2}+U(r)R_l=ER_l \qquad (35.21)$$

At large distances from the scatterer the centrifugal and potential energies tend to zero, so that the asymptotic form of (35.21) is

$$-\frac{h^2}{2m}\frac{d^2R_l}{dr^2}=ER_l \qquad (35.22)$$

The general solution to this equation is

$$R_l=A_{l\text{'}}\sin\left(kr+\delta_l-\frac{\pi l}{2}\right) \qquad (35.23)$$

Here, $k=(2mE)^{1/2}/h$. Since solution (35.23) is asymptotic, it need not become zero for $r=0$. The term $-(\pi l/2)$ is introduced to make the solution exact for a particle in free motion, that is, for $U=0$ (how this occurs will be shown later in this section). The term δ_l, called the *phase shift*, shows by how much the phase of a particle moving in the field of the scatterer differs from the phase of a free particle.

In order to compute δ_l in explicit form we must find the solution to the exact equation (35.21), which at the origin of the coordinate system vanishes as r^{l+1} (see (29.20)), and continue it into the domain where the asymptotic solution (35.23) becomes valid. We shall consider δ_l to be known, but first let us determine the condition for its existence.

At large distances from the scatterer the potential and centrifugal energies are close to zero, and the particle is almost free. Its motion in this case is quasi-classical. But in such an approximation (35.21) looks like

$$R_l\sim\sin\left\{\frac{1}{h}\int_{r_0}^{r}dr\left[2m\left(E-U(r)-\frac{h^2(l+1/2)^2}{2mr^2}\right)\right]^{1/2}+\frac{\pi}{4}\right\}$$

(r_0 is the turning point where the radicand is zero (cf. (31.40)); for a free particle, that is, for $U(r)=0$, the solution is

$$R_l^0\sim\sin\left\{\frac{1}{h}\int_{r_0}^{r}dr\left[2m\left(E-\frac{h^2(l+1/2)^2}{2mr^2}\right)\right]^{1/2}+\frac{\pi}{4}\right\}$$

It is apparent that in the quasi-classical approximation the difference between the phases of the wave functions of a free particle

and a particle moving in the field of the scatterer is

$$\delta_l \sim \frac{1}{\hbar} \left\{ \int_{r_0}^{r} dr \left[2m \left(E - U(r) - \frac{\hbar^2(l+1/2)^2}{2mr^2} \right) \right]^{1/2} \right.$$

$$\left. - \int_{r_0}^{r} dr \left[2m \left(E - \frac{\hbar^2(l+1/2)^2}{2mr^2} \right) \right]^{1/2} \right\} \qquad (35.24)$$

If it tends to a constant value when $r \to \infty$, then the phase exists always, and not only in the quasi-classical approximation, because the domain of large r's is the determining factor. Expanding the integrand in (35.24) in a series in $U(r)$, we see that δ_l is finite if the integral

$$\int_{r_0}^{\infty} \frac{(2m)^{1/2} U(r) \, dr}{[E - \hbar^2(l+1/2)^2/(2mr^2)]^{1/2}} \qquad (35.25)$$

converges, which in turn depends on the convergence of $\int_{r_0}^{\infty} dr \, U(r)$ for large r. If $U(r)$ is a power function, then for the integral $\int_{r_0}^{r} U(r) \, dr$ to converge it is sufficient for the exponent in $U(r) = a/r^n$ to be greater than unity. Thus, finite phases do not exist for the Coulomb field, but as was shown, an exact solution of the scattering problem for it is possible.

Assuming the existence of δ_l, let us find the factors in the asymptotic wave functions, A_l. We note that solution (35.23) involves the incident and scattered waves: it is a general solution. It is not hard to determine the asymptotic form of the incident wave: at a sufficient distance from the scatterer it is a plane wave. Since we suppose that the particles are travelling along the polar axis, their wave function for large distances from the scatterer must be written as e^{ikz}. Such a function is normalized to unity in unit volume (cf. (35.1)), therefore the flux density of the incident particles is represented simply as v.

The difference $\psi - e^{ikz}$ is thus an asymptotic expression of a scattered wave, provided that in place of ψ we have substituted its value for large distances from the scatterer. But at an infinite distance the scattered wave should not contain particles moving towards the scatterer. In other words, at infinity the function $\psi - e^{ikz}$ should involve only outgoing waves of the type e^{ikr}/r. (Note that the wave contains, together with time dependence, the dependence

32*

$r^{-1}e^{-iEt/h+ikr}$, which shows that the motion is in the direction of increasing r.)

The function (35.23) tnvolves both ingoing and outgoing waves. If we expand e^{ikz} in Legendre polynomials, the expansion coefficients dependent upon r will also correspond to both waves. We must so choose the constant coefficients A_l that the difference $\psi - e^{ikz}$ contain only the outgoing wave.

Let us expand $e^{ikz} = e^{ikr \cos \theta}$ in Legendre polynomials, assuming r very large, since we are concerned only with the asymptotic form of the functions of r. Let

$$e^{ikr \cos \theta} = \frac{1}{r} \sum_{l'=0}^{\infty} R_{l'}^0 (r) P_{l'} (\cos \theta) \tag{35.26}$$

We multiply both sides of (35.26) by $P_l (\cos \theta)$ and integrate over $\sin \theta \, d\theta$. In the right-hand side we obtain, from (29.10),

$$\sum_{l'} R_{l'}^0 \int_0^{\pi} P_l (\cos \theta) P_{l'} (\cos \theta) \sin \theta \, d\theta$$

$$= \sum_{l'} R_{l'}^0 \frac{2}{2l+1} \delta u'$$

$$= \frac{2R_l^0 (r)}{2l+1} \tag{35.27}$$

In the left-hand side, we integrate once by parts to get

$$-\int_0^{\pi} e^{ikr \cos \theta} P_l (\cos \theta) \, d \cos \theta$$

$$= \int_{-1}^{1} e^{ikr\xi} P_l (\xi) \, d\xi$$

$$= \frac{e^{ikr\xi}}{ikr} P_l (\xi) - \frac{1}{ikr} \int_{-1}^{1} \frac{dP_l (\xi)}{d\xi} e^{ikr\xi} \, d\xi$$

The integrated expression involves r in the denominator; the next term, if we integrate it once again, yields r^2 in the denominator; etc., up to the lth derivative of $P_l (\xi)$, which is a constant number. But we must retain only the term proportional to r^{-1}, which is of importance for the asymptotic solution. We substitute the limits into it and make use of the fact that $P_l (1) = 1$, $P_l (-1) = (-1)^l$ (see Exercise 1, Section 29).

Comparing the integrals in the left- and right-hand sides of Eq. (35.26) multiplied by $P_l (\cos \theta)$, we obtain

$$R_l^0 (r) = \frac{2l+1}{2} \frac{e^{ikr} - (-1)^l e^{-ikr}}{ik}$$

$$= \frac{2l+1}{2ik} e^{i\pi l/2} [e^{i(kr - \pi l/2)} - e^{-i(hr - \pi l/2)}] \qquad (35.28)$$

where we replaced $(-1)^l$ by $e^{i\pi l} = e^{i\pi l/2} \times e^{i\pi l/2}$. Now represent (35.23) as

$$R_l (r) = \frac{A_l}{2i} [e^{i(kr + \delta_l - \pi l/2)} - e^{-i(kr + \delta_l - \pi l/2)}] \qquad (35.29)$$

The difference $R_l - R_l^0$ should not involve an ingoing wave, hence

$$A_l = \frac{2l+1}{k} e^{i\delta_l + i\pi l/2} \qquad (35.30)$$

Substituting this value of A_l into (35.29) and forming the difference $R_l - R_l^0$ we find the outgoing wave with angular momentum l:

$$R_l - R_l^0 = \frac{2l+1}{2ik} e^{ikr} (e^{2i\delta_l} - 1) \qquad (35.31)$$

If the term $-\pi l/2$ had not been introduced into (35.23), it would have appeared in the exponent.

Finally, the scattered wave is represented in the form

$$\psi' = \frac{e^{ikr}}{r} f(\theta) \qquad (35.32)$$

where

$$f(\theta) = \frac{1}{2ik} \sum_{l=0}^{\infty} (2l + 1) (e^{2i\delta_l} - 1) P_l (\cos \theta) \qquad (35.33)$$

Let us find the expression for the differential scattering cross section. The radial component of a flux of particles through an infinitely remote sphere in a unit solid angle is (cf. (35.2))

$$\frac{h}{2mi} \left(\psi'^* \frac{\partial \psi'}{\partial r} - \psi' \frac{\partial \psi'^*}{\partial r} \right) r^2 \, d\Omega = v \, | f(\theta) |^2 \, d\Omega$$

Dividing this relation by the flux of incident particles, which is equal to v, we reduce the expression for the differential cross section to the form

$$d\sigma = | f(\theta) |^2 \, d\Omega \qquad (35.34)$$

Equations (35.33) and (35.34) are fundamental in the theory of scattering.

If all the phase shifts are small in comparison with unity, the expression for the effective cross section transforms into the Born

approximation formula, though we shall not verify this. We shall only note the following. The convergence of the series (35.33) or **of** other series developed from it requires that lim $\delta_l = 0$. Consequently, for large l the evaluations made on the basis of the Born approximation are correct.

Making use of the orthogonality of the Legendre polynomials, we now find the *total scattering cross section*

$$\sigma = \int d\sigma = \int |f(\theta)|^2 \, d\Omega$$

$$= \frac{1}{4k^2} \int d\Omega \left| \sum_l (2l+1)(e^{2i\delta_l} - 1) P_l(\cos\theta) \right|^2$$

$$= \frac{1}{4k^2} \sum_{l',l} (2l+1)(2l'+1)(e^{2i\delta_l} - 1)(e^{-2i\delta_{l'}} - 1)$$

$$\times \int P_l(\cos\theta) P_{l'}(\cos\theta) \, d\Omega$$

$$= \frac{1}{4k^2} \sum_{l',l} (2l+1)(2l'+1)(e^{2i\delta_l} - 1)(e^{-2i\delta_{l'}} - 1) \frac{2 \times 2\pi}{2l+1} \delta_{ll'}$$

$$= \frac{\pi}{k^2} \sum_l (2l+1) |e^{2i\delta_l} - 1|^2 \qquad (35.35)$$

Thus, the total cross section is separated into a sum of *partial cross sections*, each of which corresponds to the scattering of particles having a specific angular momentum ("arm" or impact parameter in classical terms):

$$\sigma = \sum_l \sigma_l$$

where

$$\sigma_l = \frac{\pi}{k^2} (2l+1)(e^{2i\delta_l} - 1)(e^{-2i\delta_l} - 1)$$

$$= \frac{\pi}{k^2} (2l+1)(2 - e^{2i\delta_l} - e^{-2i\delta_l})$$

$$= \frac{\pi}{k^2} (2l+1) \, 2(1 - \cos 2\delta_l) = \frac{4\pi}{k^2} (2l+1) \sin^2 \delta_l \qquad (35.36)$$

The **maximum** scattering cross section for a particle with a given angular **m**omentum l is thus

$$(\sigma_l)_{max} = \frac{4\pi}{k^2} (2l+1) \qquad (35.37)$$

The relation between the angular momentum and the impact parameter ("arm") is conveniently written in the quasi-classical approximation, and as was pointed out in Section 33 (cf. (33.38)), instead

of $[l\,(l\,+\,1)]^{1/2}$ we should take $l + 1/2$ for the absolute value of the angular momentum. Then

$$l + 1/2 = k\rho \tag{35.38}$$

where ρ is the impact parameter.

Thus, the partial cross section (35.36) corresponds to a ring on the impact plane. But there is also a difference, due to the wave character of motion in quantum mechanics. In the classical scattering theory (Sec. 6), the differential scattering cross section was

$$d\sigma_{\text{classical}} = 2\pi\rho\, d\rho$$

From (35.38) $\rho = (l + 1/2)/k$, and $d\rho$, which corresponds to an increase in l by unity, is $1/k$. Therefore the classical expression would have been

$$d\sigma_{\text{classical}} = \frac{\pi}{k^2}\,(2l + 1)$$

that is, one-quarter the maximum quantum cross section.

A result similar to the quantum result is obtained in wave optics in the separation of a plane wave front into circular Fresnel zones. When all the zones but one are closed by a screen, the amplitude of the wave passing through the ring aperture is twice as great as through an equal area of unobstructed wave surface. But having the strict result (35.37), there is no need to continue to develop the analogy with approximate zone construction.

The function $f\,(\theta)$ is called the *scattering amplitude*. Its component corresponding to the angular momentum l is the *partial scattering amplitude* $f_l\,(\theta)$. The partial cross section σ_l is expressed in terms of it.

Of interest is an examination of the behaviour of the scattering amplitude $f_l\,(\theta)$ when the scattering angle θ tends to zero. Remembering that $P_l\,(1) = 1$, we find

$$f_l\,(0) = \frac{1}{2ik}\,(2l + 1)\,(e^{2i\delta_l} - 1) \tag{35.39}$$

Let us find the imaginary part of this expression:

$$\text{Im}\,f_l\,(0) = -\frac{1}{2k}\,(2l + 1)\,(\cos 2\delta_l - 1) = \frac{2l + 1}{k}\sin^2\delta_l \tag{35.40}$$

Comparing it with (35.36), we obtain the relationship between the scattering amplitude for angle zero and the partial scattering cross section:

$$\text{Im}\,f_l\,(0) = \frac{k}{4\pi}\,\sigma_l \tag{35.41a}$$

Since the proportionality factor does not involve l, the relation between the total zero-angle scattering amplitude and the total

cross section is the same:

$$\operatorname{Im} f(0) = \frac{k}{4\pi} \sigma \qquad (35.41b)$$

Separation of the total cross section into partial cross sections is especially useful when one or several of the partial amplitudes are significant.

For example, in the case of short-range forces, mainly particles in the *s* state, that is, with zero angular momentum, are scattered. The wave functions of all other states vanish in the region of action of the forces. But if only one phase is other than zero ($\delta_0 \neq 0$, $\delta_{l \neq 0} = 0$), then only the *s*th scattering amplitude, which in the expansion is the factor of the zero Legendre polynomial, $P_0 = 1$, is not zero. Hence, in the case of short-range forces the scattering is spherically symmetrical. This was already discussed in Section 6. But in the classical theory there always remains a sharp maximum of the differential scattering cross section in the direction of the incident beam. In quantum theory the *s*th cross section is strictly isotropic.

EXERCISES

1. Find the scattering cross section of fast particles by hydrogen atoms in the ground state, assuming that the state does not change.

Solution. The wave function for the ground state of a hydrogen atom with $n = 1$, $l = 0$ is

$$\psi_0 = Be^{-\xi}$$

The coefficient B is found from the normalization condition

$$\left(\frac{h^2}{me^2} \right)^3 4\pi B^2 \int_0^\infty \xi^2 e^{-2\xi}\, d\xi = 1$$

whence

$$B = \pi^{-1/2} \left(\frac{me^2}{h^2} \right)^{3/2}$$

The potential energy of the charge e in the field of the atom is

$$U = -\frac{e^2}{r} + \int \frac{e^2 \psi_0^2(r')}{|\mathbf{r} - \mathbf{r}'|}\, dV'$$

The first term in $U_{\mathbf{kk}'}$ was found in the text. It is equal to

$$-\frac{\pi e^2}{k^2 \sin^2(\theta/2)}$$

We find the second term in the following way:

$$e^2 \int e^{i(\mathbf{k}-\mathbf{k}')\mathbf{r}} \, dV \int \frac{\psi_0^2 \, (r') \, dV'}{|\, \mathbf{r} - \mathbf{r}' \,|}$$

$$= e^2 \int \psi_0^2 \, (r') \, e^{i(\mathbf{k}-\mathbf{k}')\mathbf{r}'} \, dV' \int \frac{e^{i(\mathbf{k}-\mathbf{k}')(\mathbf{r}-\mathbf{r}')}}{|\, \mathbf{r} - \mathbf{r}' \,|} \, dV$$

In the last integral it is necessary to take the origin at the point $\mathbf{r}'$. Then it reduces to the same form as the preceding one:

$$\int \frac{e^{i(\mathbf{k}-\mathbf{k}')\mathbf{r}}}{r} \, dV = \frac{\pi}{k^2 \sin^2(\theta/2)}$$

Hence

$$U_{\mathbf{k}\mathbf{k}'} = -\frac{\pi e^2}{k^2 \sin^2(\theta/2)} \left(1 - \int \psi_0^2 \, (r') \, e^{i(\mathbf{k}-\mathbf{k}')\mathbf{r}'} \, dV' \right)$$

The quantity inside the brackets is called the *screening factor*. Evaluating it in the same manner as (35.8), we obtain

$$\int \psi_0^2 \, (r') \, e^{i(\mathbf{k}-\mathbf{k}')\mathbf{r}'} \, dV' = \frac{2\pi}{k \sin\theta} \int_0^\infty r' \psi_0^2 \, (r') \sin\left[2kr' \sin(\theta/2)\right] dr'$$

The integral reduces to the form

$$\int_0^\infty x \sin ax \, e^{-bx} \, dx = -\frac{\partial}{\partial b} \int_0^\infty \sin ax \, e^{-bx} \, dx$$

$$= -\frac{\partial}{\partial b} \frac{a}{a^2 + b^2} = \frac{2ab}{(a^2 + b^2)^2}$$

Here $a = 2k \sin(\theta/2)$, $b = 2me^2/h^2$, so that the screening factor is

$$1 - \left[1 + \left(\frac{hv}{e^2} \right)^2 \sin^2 \frac{\theta}{2} \right]^{-2}$$

It was assumed in the last formula that the scattered particle is an electron, that is, it has the same mass and charge. Strictly speaking, we should have formed a function which is antisymmetric to the function of the atomic electron; this we did not do. The final formula for the elastic scattering cross section differs from (35.16) by the square of the screening factor. We note that this factor is correctly obtained only in the Born approximation, in contrast to the Rutherford formula, which is exact.

For $\theta = 0$, the cross section turns out to be finite, because $\theta = 0$ corresponds to large impact parameters when the nuclear charge is screened by the charge of an electron.

2. Calculate the effective scattering cross section for a particle scattered on an opaque sphere of radius a much smaller than $\lambda/(2\pi) \equiv \lambdabar = 1/k$, so that only s scattering is significant.

Solution. From the boundary condition (28.1), on the surface of an opaque sphere the wave function vanishes. Hence, solution (35.23) has the

form

$$R_0 = A_0 \sin k\,(r - a)$$

From this $\delta_0 = -ka$, and

$$\sigma = \frac{4\pi}{k^2} \sin^2 ka$$

But by definition $ka \ll 1$, so that $\sin ka \approx ka$, and $\sigma = 4\pi a^2$, that is, the scattering radius is twice the radius of the sphere. The cross section in the quantum theory is four times larger than in the classical theory (see Exercise 1, Section 6).

3. Find the scattering cross section for particles scattered by a potential well of depth close to U_{cr}, defined by Eq. (28.44). The first bound state appears in the well at $U = U_{cr}$. It is assumed that the energy of the incident particle is much smaller than the difference $|U_{cr} - U_0|$, and the radius of the well is much smaller than the wavelength of the incident particle divided by 2π. Express the cross section in terms of the energy of the bound state.

Solution. Displace the potential energy curve (Section 28, Figure 31) along the vertical axis so as to obtain $U = 0$ for $r > a$, that is, at infinite distance from the origin we put $U\,(r)$ zero. The condition for matching the wave function at $r = a$ has the form

$$\frac{k \cos\,(ka + \delta_0)}{\sin\,(ka + \delta_0)} = \frac{\varkappa \cos \varkappa a}{\sin \varkappa a}$$

where

$$k = \frac{1}{\hbar}\,(2mE)^{1/2}, \qquad \varkappa = \frac{1}{\hbar}\,[2m\,(E + |U_0|)]^{1/2}$$

Neglecting ka, we obtain the expression for the scattering cross section:

$$\sigma = \frac{4\pi}{k^2} \sin^2 \delta_0 = \frac{4\pi}{k^2\,(1 + \cot^2 \delta_0)} = \frac{4\pi}{k^2 + \varkappa^2 \cot^2 \varkappa a}$$

Further, using the notation introduced in Section 28, we find the approximate expression for $\varkappa$:

$$\varkappa \approx \frac{(2m\,|U_{cr}|)^{1/2}}{\hbar} \left(1 + \frac{v}{2} + \frac{E}{2U_{cr}}\right)$$

$$= \frac{(2m\,|U_{cr}|)^{1/2}}{\hbar} \left(1 + \frac{U_0 - U_{cr}}{2U_{cr}} + \frac{E}{2U_{cr}}\right)$$

By definition, $U_0 - U_{cr} \gg E$, so that

$$\varkappa a \approx \frac{\pi}{2} \left(1 + \frac{v}{2}\right), \qquad \varkappa \cot \varkappa a \approx -\frac{\pi}{2a} \times \frac{\pi v}{4} = -\frac{\pi^2}{8a}\,v$$

From this

$$\varkappa^2 \cot^2 \varkappa a = \frac{\pi^4}{64a^2}\,\frac{(U_0 - U_{cr})^2}{U_{cr}^2} = \frac{2m\,|\varepsilon|}{\hbar^2} \equiv k_0^2$$

Finally, we obtain the scattering cross section in the form

$$\sigma = \frac{4\pi}{k^2 + k_0^2}$$

Note that nowhere in developing this equation did we make use of the fact that $|U_0| < |U_{cr}|$. This formula also holds for $|U_0| > |U_{cr}|$, when there is no level in the well, but it would have appeared for a slight deepening of the well. In such cases we speak of "virtual levels". Such a case of virtual level is observed in the scattering of neutrons on protons with antiparallel spins, which is established by comparing the scattering cross section of the neutrons on ortho- and parahydrogen.

4. Express the cross section in terms of the scattering amplitude for like particles mutually scattered on one another. Consider the case of particles with zero and 1/2 spin.

Solution. Consider the motion in the particles' centre-of-mass reference frame. In the case of zero spin, the wave function must be symmetrical with respect to an exchange of the particles corresponding to a substitution of $-\theta$ for θ. Hence

$$d\sigma = |f(\theta) + f(-\theta)|^2 d\Omega$$

The normalization factor $1/\sqrt{2}$ is not required here, since there are two particles and they are indistinguishable in the reaction.

In the case of 1/2 spin we have three ortho-states and one para-state. Hence, the probability of a collision in the ortho-state is 3/4, in the para-state it is 1/4, and the effective cross section is

$$d\sigma = \left\{ \frac{3}{4}|f(\theta) - f(-\theta)|^2 + \frac{1}{4}|f(\theta) + f(-\theta)|^2 \right\} d\Omega$$

$$= \left[|f(\theta)|^2 + |f(-\theta)|^2 - \frac{1}{2}f(\theta)f^*(-\theta) - \frac{1}{2}f^*(\theta)f(-\theta) \right] d\Omega$$

In both cases there appears an "interference" term ff^*, which does not appear in the analogous classical problem (cf. (6.22)). In the limiting transition to classical mechanics the interference term does not tend to zero, because it retains the final value of the scattering amplitude f: σ is expressed in terms of it. Since in the limit the wavelength λ tends to zero, ff^* becomes a very rapidly oscillating function, the mean of which over an infinitesimal interval of angles tends to zero. The limiting transition from wave optics to geometrical optics at the boundary of a shadow takes place in approximately the same way.

36

THE QUANTUM THEORY OF RADIATION

An electromagnetic field in vacuum can be treated as a mechanical system (this was shown in Section 15). It is characterized by a Lagrangian, action, etc. It is therefore legitimate to pose the problem of quantizing this system, that is, applying quantum mechanics to it.

The basic distinction between electrodynamics and the mechanics of mass points is that the degrees of freedom of an electromagnetic field are distributed continuously: to define a field at a given instant, its value at every spatial point must be defined. In this sense electrodynamics resembles the mechanics of a fluid or an elastic body, if they are treated as continuous media, disregarding the atomic structure of matter. The coordinates of spatial points as it were number the degrees of freedom of the field, while the values of the potential amplitudes define the generalized coordinates.

The electromagnetic field coordinates thus defined are not mutually independent. Indeed, the electromagnetic field equations involve derivatives with respect to the coordinates, that is, differences between values of the field at infinitely close points. In this sense the field equations resemble the equations for coupled oscillations: they are linear and involve several generalized coordinates taken for infinitely close points in space. The equations of coupled oscillations reduce to normal coordinates, which are mutually independent (Sec. 7). The same can be done with the equations of electrodynamics, thereby separating the dependent variables. This greatly simplifies the application of quantum mechanics to radiation.

Here we find vivid proof of the generality of the methods of analytical mechanics: they make it possible to so define the generalized coordinates and momenta as to make it possible to apply the quantum laws uniquely.

Electromagnetic Field in a Closed Volume. First of all, we must visualize an electromagnetic field as a closed system, since quantum mechanics is most conveniently applied precisely to such systems. We may assume, for example, that an electromagnetic field is contained in a box with mirror reflecting walls. The normal components of the Poynting vector vanish on the walls of such an imaginary box ($x = 0$ or $x = a_1$, $y = 0$ or $y = a_2$, $z = 0$ or $z = a_3$).

Let us fill all space with such boxes and assume that the field has the same value at corresponding points in each box. Such a field

is periodic in all three spatial directions:

$$\mathbf{A}(x, y, z) = \mathbf{A}(x + a_1, y, z)$$
$$= \mathbf{A}(x, y + a_2, z) = \mathbf{A}(x, y, z + a_3) \qquad (36.1)$$

But if we remove the reflecting walls, the field remains periodic anyway, because all its values are displaced at all points of space with the same fundamental velocity c. It is therefore sufficient to impose periodicity conditions (36.1) on the field and simply discard the walls. This appreciably simplifies the calculations, and the final results cannot depend on this or that supplementary device.

A solution of the equations describing an electromagnetic field in vacuum was found in Section 18. Since the periodicity condition is imposed on the field, it can be expanded in a Fourier series in all three dimensions, that is, it can be represented in terms of separate harmonic components. These components must be real quantities. Expressing them explicitly only as functions of the coordinates, we write, by analogy with (18.25),

$$\mathbf{A}(\mathbf{k, r}) = \mathbf{A_k}e^{i\mathbf{kr}} + \mathbf{A_k^*}e^{-i\mathbf{kr}} \qquad (36.2)$$

for one harmonic component. The reality of the vector potential $\mathbf{A}$ is apparent from this notation.

Since we are considering the field in vacuum (in the absence of charges), the scalar potential can be put equal to zero, and the Lorentz condition (12.42) for the vector potential is reduced to the form div $\mathbf{A} = 0$, that is (see (11.27))

$$\text{div } \mathbf{A}(\mathbf{k, r}) = \text{div}(\mathbf{A_k}e^{i\mathbf{kr}}) + \text{div}(\mathbf{A_k^*}e^{-i\mathbf{kr}})$$
$$= (\mathbf{A_k} \cdot \text{grad } e^{i\mathbf{kr}}) + (\mathbf{A_k^*} \cdot \text{grad } e^{-i\mathbf{kr}})$$
$$= i(\mathbf{k} \cdot \mathbf{A_k})e^{i\mathbf{kr}} - i(\mathbf{k} \cdot \mathbf{A_k^*})e^{-i\mathbf{kr}} = 0$$

For this equation to hold for all $\mathbf{r}$'s, the coefficient of each exponent must be equal to zero. In other words, vectors $\mathbf{A_k}$ and $\mathbf{A_k^*}$ are perpendicular to the wave vector $\mathbf{k}$:

$$(\mathbf{k} \cdot \mathbf{A_k}) = 0, \qquad (\mathbf{k} \cdot \mathbf{A_k^*}) = 0 \qquad (36.3)$$

For every $\mathbf{k}$ there exist two mutually perpendicular vectors $\mathbf{A_k^\sigma}$ ($\sigma = 1, 2$) corresponding to two possible polarizations of the wave. It is natural to choose $\mathbf{A_k^{(1)}}$ and $\mathbf{A_k^{(2)}}$ mutually perpendicular. Any vector in a plane perpendicular to $\mathbf{k}$ can be resolved along $\mathbf{A_k^{(1)}}$ and $\mathbf{A_k^{(2)}}$.

Let us now apply the periodicity condition (36.1) separately to each component of (36.2). We obtain

$$\mathbf{A_k}e^{i(k_x x + k_y y + k_z z)} = \mathbf{A_k}e^{i[k_x(x+a_1) + k_y y + k_z z]}$$
$$= \mathbf{A_k}e^{i[k_x x + k_y(y+a_2) + k_z z]} = \mathbf{A_k}e^{i[k_x x + k_y y + k_z(z+a_3)]}$$

whence it follows that

$$e^{ik_x a_1} = e^{ik_y a_2} = e^{ik_z a_3} = 1$$

Therefore, the wave vector components are

$$k_x = \frac{2\pi n_1}{a_1}, \qquad k_y = \frac{2\pi n_2}{a_2}, \qquad k_z = \frac{2\pi n_3}{a_3} \tag{36.4}$$

with n_1, n_2, and n_3 being integers of any sign.

It follows that every harmonic oscillation is given by three integers n_1, n_2, and n_3, and its polarization σ, which takes on two values. The generalized coordinate is, as stated before, $A^\sigma_{n_1 n_2 n_3}$. The number of such coordinates is infinite, but it at least constitutes a countable set, not a continuous assembly similar to the assembly of all spatial points.

This is the basic simplification introduced by the periodicity condition. Of course, it pursues only mathematical convenience: the basic periods a_1, a_2, a_3 are not involved in any final results.

An electromagnetic field is defined if the amplitudes of its oscillations for all values of n_1, n_2, n_3, and σ are known. Since the equations of electrodynamics are linear, their general solution is equal to the sum of the partial solutions (36.2):

$$\mathbf{A}(\mathbf{r}) = \sum_{\mathbf{k},\,\sigma} \mathbf{A}^\sigma_\mathbf{k}(\mathbf{r}) = \sum_{\mathbf{k},\,\sigma} (A^\sigma_\mathbf{k} e^{i\mathbf{kr}} + A^{\sigma*}_\mathbf{k} e^{-i\mathbf{kr}}) \tag{36.5}$$

This is the required Fourier expansion.

Energy of Field. We shall show that the expansion (36.5) provides all that is necessary to develop the normal coordinates. For this we must express the energy of the field in terms of $A^\sigma_\mathbf{k}$. An electric field is calculated according to the general equation (12.35), which for $\varphi = 0$ yields

$$\mathbf{E} = -\frac{1}{c}\frac{\partial \mathbf{A}}{\partial t} = -\frac{1}{c}\sum_{\mathbf{k},\,\sigma}(\dot{A}^\sigma_\mathbf{k} e^{i\mathbf{kr}} + \dot{A}^{\sigma*}_\mathbf{k} e^{-i\mathbf{kr}}) \tag{36.6}$$

and for a magnetic field

$$\mathbf{H} = \operatorname{curl}\mathbf{A} = \sum_{\mathbf{k},\,\sigma}[(\operatorname{grad} e^{i\mathbf{kr}} \times A^\sigma_\mathbf{k}) + (\operatorname{grad} e^{-i\mathbf{kr}} \times A^\sigma_\mathbf{k})]$$

$$= i\sum_{\mathbf{k},\,\sigma}[(\mathbf{k} \times A^\sigma_\mathbf{k}) e^{i\mathbf{kr}} - (\mathbf{k} \times A^{\sigma*}_\mathbf{k}) e^{-i\mathbf{kr}}] \tag{36.7}$$

Let us now calculate the field's energy, which from (15.24c) is

$$E = \frac{1}{8\pi}\int (|\mathbf{E}|^2 + |\mathbf{H}|^2)\, dV \tag{36.8}$$

For $|\mathbf{E}|^2$ we have the following sum over $\mathbf{k}$, $\mathbf{k'}$, σ, σ':

$$|\mathbf{E}|^2 = \sum_{\mathbf{k},\,\mathbf{k'},\,\sigma,\,\sigma'} \frac{1}{c^2} \left(\dot{A}_{\mathbf{k}}^{\sigma} \dot{A}_{\mathbf{k'}}^{\sigma'} e^{i(\mathbf{k}+\mathbf{k'})\mathbf{r}} + \dot{A}_{\mathbf{k}}^{\sigma} \dot{A}_{\mathbf{k'}}^{\sigma'*} e^{i(\mathbf{k}-\mathbf{k'})\mathbf{r}} \right.$$

$$\left. + \dot{A}_{\mathbf{k}}^{\sigma*} \dot{A}_{\mathbf{k'}}^{\sigma'} e^{-i(\mathbf{k}-\mathbf{k'})\mathbf{r}} + \dot{A}_{\mathbf{k}}^{\sigma*} \dot{A}_{\mathbf{k'}}^{\sigma'} e^{-i(\mathbf{k}+\mathbf{k'})\mathbf{r}} \right) \quad (36.9)$$

Integrating $|\mathbf{E}|^2$ over the volume, it is useful to put the integral under the summation sign. Then each integral separates into a product of three integrals of the following form:

$$\int_0^{a_1} e^{i(h_x + h'_x)x}\,dx = \int_0^{a_1} e^{2\pi i x(n_1 + n'_1)/a_1}\,dx$$

$$= \frac{a_1 \left[e^{2\pi i (n_1 + n'_1)} - 1 \right]}{2\pi i\,(n_1 + n'_1)} = 0 \quad (36.10)$$

If $n_1 + n'_1 = 0$, this integral is equal to a_1. Therefore, the triple integral reduces to the following expression:

$$\int e^{i(\mathbf{k}+\mathbf{k'})\mathbf{r}}\,dV = a_1 a_2 a_3 \delta_{n_1 - n'_1} \delta_{n_2 - n'_2} \delta_{n_3 - n'_3} = V\delta_{\mathbf{k}-\mathbf{k'}} \quad (36.11)$$

Hence, the double sum over $\mathbf{k}$ and $\mathbf{k'}$ in $\int |\mathbf{E}|^2\,dV$ reduces to a single sum; in the terms involving the product $\dot{A}_{\mathbf{k}}^{\sigma} \dot{A}_{\mathbf{k'}}^{\sigma'}$ we should replace $\mathbf{k'}$ by $-\mathbf{k}$, and where we have $\dot{A}_{\mathbf{k}}^{\sigma} \dot{A}_{\mathbf{k'}}^{\sigma'*}$, $\mathbf{k'}$ must be replaced simply by $\mathbf{k}$, because the factor of $\dot{A}_{\mathbf{k'}}^{\sigma'*}$ is $e^{-i\mathbf{k'}\mathbf{r}}$. Thus

$$\int |\mathbf{E}|^2\,dV = \frac{V}{c^2} \sum_{\mathbf{k},\,\sigma,\,\sigma'} \left(\dot{A}_{\mathbf{k}}^{\sigma} \dot{A}_{\mathbf{k}}^{\sigma'*} + \dot{A}_{\mathbf{k}}^{\sigma*} \dot{A}_{\mathbf{k}}^{\sigma'} \right.$$

$$\left. + \dot{A}_{\mathbf{k}}^{\sigma} \dot{A}_{-\mathbf{k}}^{\sigma'} + \dot{A}_{\mathbf{k}}^{\sigma*} \dot{A}_{-\mathbf{k}}^{\sigma'*} \right) \quad (36.12)$$

But the vectors $\dot{A}_{\mathbf{k}}^{\sigma}$ and $\dot{A}_{-\mathbf{k}}^{\sigma'}$, $\dot{A}_{\mathbf{k}}^{\sigma}$ and $A_{\mathbf{k}}^{\sigma'*}$ are perpendicular, if $\sigma \neq \sigma'$. Hence, instead of the double sum over σ and σ' there also remains the single sum over σ:

$$\int |\mathbf{E}|^2\,dV = \frac{V}{c^2} \sum_{\mathbf{k},\,\sigma} \left(2\dot{A}_{\mathbf{k}}^{\sigma} \dot{A}_{\mathbf{k}}^{\sigma*} + \dot{A}_{\mathbf{k}}^{\sigma} \dot{A}_{-\mathbf{k}}^{\sigma} + \dot{A}_{\mathbf{k}}^{\sigma*} \dot{A}_{-\mathbf{k}}^{\sigma*} \right) \quad (36.13)$$

Equation (36.11) should also be used to calculate the integral of the square of the magnetic field. But if $\mathbf{k'} = -\mathbf{k}$, the product $\mathbf{k'} \times A_{-\mathbf{k'}}^{\sigma'}$ is replaced by $-\mathbf{k} \times A_{-\mathbf{k}}^{\sigma'}$. Therefore

$$\int |\mathbf{H}|^2\,dV = V \sum \left[(\mathbf{k} \times A_{\mathbf{k}}^{\sigma})(\mathbf{k} \times A_{-\mathbf{k}}^{\sigma}) + (\mathbf{k} \times A_{\mathbf{k}}^{\sigma*})(\mathbf{k} \times A_{-\mathbf{k}}^{\sigma'*}) \right.$$

$$\left. + 2(\mathbf{k} \times A_{\mathbf{k}}^{\sigma})(\mathbf{k} \times A_{\mathbf{k}}^{\sigma'*}) \right] \quad (36.14)$$

The vector products are expressed according to the known formulas:

$$(\mathbf{k} \times \mathbf{A}_{\mathbf{k}}^{\sigma})(\mathbf{k} \times \mathbf{A}_{\mathbf{k}}^{\sigma'*}) = k^2 \mathbf{A}_{\mathbf{k}}^{\sigma} \mathbf{A}_{\mathbf{k}}^{\sigma'*} - (\mathbf{k} \cdot \mathbf{A}_{\mathbf{k}}^{\sigma})(\mathbf{k} \cdot \mathbf{A}_{\mathbf{k}}^{\sigma'*})$$

$$= k^2 \mathbf{A}_{\mathbf{k}}^{\sigma} \mathbf{A}_{\mathbf{k}}^{\sigma*} \tag{36.15}$$

where the transverse condition (36.3) has been used. For $\sigma \neq \sigma'$ the expression (36.15) becomes zero. Consequently

$$\int |\mathbf{H}|^2 dV = V \sum_{\mathbf{k}, \sigma} k^2 (2\mathbf{A}_{\mathbf{k}}^{\sigma} \mathbf{A}_{\mathbf{k}}^{\sigma*} + \mathbf{A}_{\mathbf{k}}^{\sigma*} \mathbf{A}_{-\mathbf{k}}^{\sigma*} + \mathbf{A}_{\mathbf{k}}^{\sigma} \mathbf{A}_{-\mathbf{k}}^{\sigma}) \tag{36.16}$$

Let us show that if $\mathbf{A}_{\mathbf{k}}^{\sigma}$ and $\mathbf{A}_{\mathbf{k}}^{\sigma*}$ are considered to be quantum operators and expressed in terms of the operators of position and momentum of a linear harmonic oscillator according to the formulas

$$\hat{\mathbf{A}}_{\mathbf{k}}^{\sigma} = \left(\frac{\pi c^2}{V}\right)^{1/2} \left(\hat{Q}_{\mathbf{k}}^{\sigma} + \frac{i\hat{P}_{\mathbf{k}}^{\sigma}}{\omega_{\mathbf{k}}}\right) \mathbf{e}_{\mathbf{k}}^{\sigma} \tag{36.17a}$$

$$\hat{\mathbf{A}}_{\mathbf{k}}^{\sigma*} = \left(\frac{\pi c^2}{V}\right)^{1/2} \left(\hat{Q}_{\mathbf{k}}^{\sigma} - \frac{i\hat{P}_{\mathbf{k}}^{\sigma}}{\omega_{\mathbf{k}}}\right) \mathbf{e}_{\mathbf{k}}^{\sigma} \tag{36.17b}$$

then the electromagnetic field energy (36.8) is reduced to the sum of the energy operators of linear independent harmonic oscillators. The quantity $\mathbf{e}_{\mathbf{k}}^{\sigma}$ in equations (36.17a) and (36.17b) is the *unit polarization vector of the electromagnetic wave*; $\omega\mathbf{k} = ck$.

If $\hat{Q}_{\mathbf{k}}^{\sigma}$ and $\hat{P}_{\mathbf{k}}^{\sigma}$ are the operators of position and momentum of a linear harmonic oscillator, they satisfy the quantum equations of motion (27.18) and (27.19), in which we put $m = 1$:

$$\dot{\hat{Q}}_{\mathbf{k}}^{\sigma} = \hat{P}_{\mathbf{k}}^{\sigma} \tag{36.18}$$

$$\dot{\hat{P}}_{\mathbf{k}}^{\sigma} = -\omega_{\mathbf{k}}^2 \hat{Q}_{\mathbf{k}}^{\sigma} \tag{36.19a}$$

Then from equations (36.17a) and (36.17b) it follows that

$$\dot{\hat{\mathbf{A}}}_{\mathbf{k}}^{\sigma} = \left(\frac{\pi c^2}{V}\right)^{1/2} \left(\dot{\hat{Q}}_{\mathbf{k}}^{\sigma} + \frac{i\dot{\hat{P}}_{\mathbf{k}}^{\sigma}}{\omega_{\mathbf{k}}}\right) \mathbf{e}_{\mathbf{k}}^{\sigma}$$

$$= i\omega_{\mathbf{k}} \left(\frac{\pi c^2}{V}\right)^{1/2} \left(\hat{Q}_{\mathbf{k}}^{\sigma} + \frac{i\hat{P}_{\mathbf{k}}^{\sigma}}{\omega_{\mathbf{k}}}\right) \mathbf{e}_{\mathbf{k}}^{\sigma} = -i\omega_{\mathbf{k}} \hat{\mathbf{A}}_{\mathbf{k}}^{\sigma} \tag{36.19b}$$

$$\dot{\hat{\mathbf{A}}}_{\mathbf{k}}^{\sigma*} = \left(\frac{\pi c^2}{V}\right)^{1/2} \left(\dot{\hat{Q}}_{\mathbf{k}}^{\sigma} - \frac{i\dot{\hat{P}}_{\mathbf{k}}^{\sigma}}{\omega_{\mathbf{k}}}\right) \mathbf{e}_{\mathbf{k}}^{\sigma}$$

$$= i\omega_{\mathbf{k}} \left(\frac{\pi c^2}{V}\right)^{1/2} \left(\hat{Q}_{\mathbf{k}}^{\sigma} - \frac{i\hat{P}_{\mathbf{k}}^{\sigma}}{\omega_{\mathbf{k}}}\right) \mathbf{e}_{\mathbf{k}}^{\sigma} = i\omega_{\mathbf{k}} \hat{\mathbf{A}}_{\mathbf{k}}^{\sigma} \tag{36.19c}$$

If we substitute these expressions into (36.13), the last two terms in the parentheses yield

$$\hat{\dot{A}}_{\mathbf{k}}^{\sigma}\hat{\dot{A}}_{-\mathbf{k}}^{\sigma} + \hat{\dot{A}}_{\mathbf{k}}^{\sigma*}\hat{\dot{A}}_{-\mathbf{k}}^{\sigma*} = -\omega_{\mathbf{k}}^2(\hat{A}_{\mathbf{k}}^{\sigma}\hat{A}_{-\mathbf{k}}^{\sigma} + \hat{A}_{\mathbf{k}}^{\sigma*}\hat{A}_{-\mathbf{k}}^{\sigma*})$$

When substituted into the total energy formula, these two terms from the integral $\int |\mathbf{E}|^2\, dV$ cancel out with the last two terms from $\int |\mathbf{H}|^2\, dV$, that is, from (36.14). The operators $\hat{A}_{\mathbf{k}}^{\sigma}$ and $\hat{A}_{\mathbf{k}}^{\sigma*}$ do not commute, so to pass from the classical to the quantum expression for energy we must replace $\hat{A}_{\mathbf{k}}^{\sigma}\hat{A}_{\mathbf{k}}^{\sigma*}$ by $(\hat{A}_{\mathbf{k}}^{\sigma}\hat{A}_{\mathbf{k}}^{\sigma*} + \hat{A}_{\mathbf{k}}^{\sigma*}\hat{A}_{\mathbf{k}}^{\sigma})/2$. Substituting (36.17a) and (36.17b) into it, we obtain

$$\frac{1}{8\pi} V\left(\frac{\hat{\dot{A}}_{\mathbf{k}}^{\sigma}\hat{\dot{A}}_{\mathbf{k}}^{\sigma*} + \hat{\dot{A}}_{\mathbf{k}}^{\sigma*}\hat{\dot{A}}_{\mathbf{k}}^{\sigma}}{2c^2} + k^2\frac{\hat{A}_{\mathbf{k}}^{\sigma}\hat{A}_{\mathbf{k}}^{\sigma*} + \hat{A}_{\mathbf{k}}^{\sigma*}\hat{A}_{\mathbf{k}}^{\sigma}}{2}\right)$$
$$= \frac{(\hat{P}_{\mathbf{k}}^{\sigma})^2 + \omega_{\mathbf{k}}^2(\hat{Q}_{\mathbf{k}}^{\sigma})^2}{2} \qquad (36.20)$$

that is, the Hamiltonian of a linear harmonic oscillator of unit mass.

If we do not make the product $\hat{A}_{\mathbf{k}}^{\sigma}\hat{A}_{\mathbf{k}}^{\sigma*}$ symmetrical, the energy acquires a constant term which is immaterial for the subsequent discourse. But the form (36.20) is the standard one for the Hamiltonian of an oscillator. By reducing the Hamiltonian to such a form we have justified equations (36.18) and (36.19a), which are derived from the Hamiltonian (36.20) as quantum equations of motion.

Quanta. Thus, the Hamiltonian of an electromagnetic field in which there are no charges is expressed as the sum of the Hamiltonians of linear independent harmonic oscillators, each of which corresponds to one value of the wave vector $\mathbf{k}$ and to one polarization σ. To these oscillators can be applied all the quantization rules obtained in Section 27. In other words, they can be described in the representation in which the energy of each oscillator is diagonal.

The energy eigenvalue of a separate oscillator is determined from Eqs. (27.23a) and (27.28b):

$$E_{\mathbf{k}}^{\sigma} = \hbar\omega_{\mathbf{k}}\left(N_{\mathbf{k}\sigma} + \frac{1}{2}\right) \qquad (36.21)$$

Here, the term $\hbar\omega_{\mathbf{k}}/2$ corresponds to the ground state of the oscillator, the number $N_{\mathbf{k}\sigma}$ gives the number of quanta of frequency ω, wave vector $\mathbf{k}$, and polarization σ in the field.

The momentum of an electromagnetic field can be found (Exercise 2) similarly to the way the energy was calculated. It then turns

out that to the oscillator $\mathbf{k}$, σ there corresponds the momentum $\mathbf{p_k^\sigma} = \hbar\mathbf{k}\ (N_{\mathbf{k}\sigma} + 1/2)$, confirming the ratio between the energy and momentum of a quantum adopted before. Thus, quanta are not an additional hypothesis concerning the properties of an electromagnetic field, but a corollary of the quantum laws of motion as applied to fields. It is wrong to imagine that quanta "explain" the nature of light. One could claim with equal justification that the formula $E_n = \hbar\omega\ (n + 1/2)$ explains the nature of vibrational motion in general.

Quanta should certainly not be treated as the fruit of sundry mathematical exercises that result in Eq. (36.21). A quantum is a real, existing particle, the same as an electron. For example, in the scattering of X-ray quanta on electrons, the energy $\hbar\omega$ and momentum $\hbar\mathbf{k}$ of each separate quantum are involved in the general law of conservation of energy and momentum in collisions in the same way as for any other particle. In scattering the frequency of a quantum decreases in proportion to its energy. Like the electron, which possesses an intrinsic degree of freedom (spin), the quantum possesses a polarization degree of freedom. But it cannot be identified with 1/2 spin, because the quantum is described with the help of a vector quantity—the vector potential—while 1/2 spin is described by spinors (Sec. 30).

The latter is due to the fact that the limiting transition to the classical theory for the quantum and for the electron occurs in entirely different ways. There is nothing in the limiting transition to classical mechanics that corresponds to the quantum: its energy and momentum, $\hbar\omega$ and $\hbar\mathbf{k}$, tend to zero when $\hbar \to 0$. The situation with the electron is different, and its momentum and energy pass from quantum to classical quantities.

Matters are different again with the wave properties of motion. In the limiting transition to the classical theory, the energy of each quantum is considered to be infinitely small, and their number $N_{\mathbf{k}\sigma}$ infinitely large, so that the wave amplitude remains finite.

Because of their half-integral spin, electrons are subject to Pauli's exclusion principle: no two electrons can be in the same state. Accordingly, in the classical transition there is nothing that corresponds to the wave function of the electron (and hence to all the wave properties of its motion). They appear only in quantum theory.

It should be noted that the amplitude of an electromagnetic wave should not be identified with the wave function of a quantum, interpreting it in the same sense as the probability amplitude of an electron. The square of the wave amplitude is used to express the energy density of a field, not the quantum density. If we wished to pass to quantum density, the quantity would have to be divided by the corresponding value of ω for each frequency. This would give the expression for quantum density in the momentum representation.

But in the coordinate representation, which is obtained from the momentum representation by means of a Fourier transformation, quantum density cannot be expressed in terms of the square of the wave amplitude or its derivatives, because in passing to the coordinate representation the frequency division operator does not yield the δ function or its derivatives.

Let us now consider the state of a field when all $N_{\mathbf{k}\sigma} = 0$. Apparently, this is the ground state, which is also conventionally called a *vacuum* (with respect to the quanta). In it all the separate oscillators used to represent the field are in the ground state. But as was shown in Section 28, the coordinate of a ground-state oscillator is not zero: it does not have a strictly defined value. The probability of a certain value of the coordinate, $Q_{\mathbf{k}}^{\sigma}$, is described by the square of the oscillator wave function, $\psi_0^2(Q_{\mathbf{k}}^{\sigma})$. (This wave function has no relation to quanta: it states the probability of certain field values!)

Thus, even in the ground state (vacuum), in the absence of charges, an electromagnetic field does not vanish. Its amplitude **A** is expressed in terms of the oscillator coordinates $Q_{\mathbf{k}}^{\sigma}$. This leads to observable effects, which will be discussed in the next section.

Interaction of an Electromagnetic Field with a Charged Particle. We shall now examine radiation transitions, that is, processes of quantum emission and absorption in interactions with charged particles. For that we must find the operator describing the interaction. The corresponding classical quantity for a separate charge is obtained from (17.32) for $\varphi = 0$:

$$\mathcal{H}' = -\frac{e}{mc}(\mathbf{p}\cdot\mathbf{A}) \tag{36.22}$$

It corresponds to a nonrelativistic approximation with respect to a moving charge: its velocity is assumed to be small in comparison with the speed of light.

To pass on to operators we must, as usual, replace **p** by $(h/i)\nabla$ and substitute in place of the vector potential the operator expression corresponding to (36.5), replacing in it the amplitudes of separate harmonic waves by the operators (36.19b) and (36.19c). We agreed to represent these operators in such a way as to have all the numbers of quanta in all states diagonal. For that we must substitute instead of the operators $\hat{Q}_{\mathbf{k}}^{\sigma}$ and $\hat{P}_{\mathbf{k}}^{\sigma}$ their matrix expressions (27.28a).

We substitute the operators (36.19b) and (36.19c) into the vector potential expansion (36.5), assuming for the sake of simplicity that the volume is equal to unity, since it is in any case eliminated

from the final formulas:

$$\hat{\mathbf{A}} = \sum_{\mathbf{k},\,\sigma} (\pi c^2)^{1/2}\, \mathbf{e}_{\mathbf{k}}^{\sigma} \left[\left(\hat{Q}_{\mathbf{k}}^{\sigma} + \frac{i\hat{P}_{\mathbf{k}}^{\sigma}}{\omega_{\mathbf{k}}} \right) e^{i\mathbf{k}\mathbf{r}} \right.$$

$$\left. + \left(\hat{Q}_{\mathbf{k}}^{\sigma} - \frac{i\hat{P}_{\mathbf{k}}^{\sigma}}{\omega_{\mathbf{k}}} \right) e^{-i\mathbf{k}\mathbf{r}} \right] \qquad (36.23)$$

Consequently, we have to develop the matrices of $\hat{Q}_{\mathbf{k}}^{\sigma} + i\hat{P}_{\mathbf{k}}^{\sigma}/\omega_{\mathbf{k}}$ and $\hat{Q}_{\mathbf{k}}^{\sigma} - i\hat{P}_{\mathbf{k}}^{\sigma}/\omega_{\mathbf{k}}$.

With the help of the matrices (27.28a) we obtain

$$\hat{Q}_{\mathbf{k}}^{\sigma} + \frac{i\hat{P}_{\mathbf{k}}^{\sigma}}{\omega_{\mathbf{k}}} = \left(\frac{2h}{\omega_{\mathbf{k}}} \right)^{1/2} \begin{pmatrix} 0 & 1 & 0 & 0 & \dots \\ 0 & 0 & \sqrt{2} & 0 & \dots \\ 0 & 0 & 0 & \sqrt{3} & \dots \\ 0 & 0 & 0 & 0 & \dots \\ \multicolumn{5}{c}{\dotfill} \end{pmatrix}$$

$$\hat{Q}_{\mathbf{k}}^{\sigma} - \frac{i\hat{P}_{\mathbf{k}}^{\sigma}}{\omega_{\mathbf{k}}} = \left(\frac{2h}{\omega_{\mathbf{k}}} \right)^{1/2} \begin{pmatrix} 0 & 0 & 0 & 0 & \dots \\ 1 & 0 & 0 & 0 & \dots \\ 0 & \sqrt{2} & 0 & 0 & \dots \\ 0 & 0 & \sqrt{3} & 0 & \dots \\ \multicolumn{5}{c}{\dotfill} \end{pmatrix} \qquad (36.24)$$

It is convenient to separate out the matrices themselves, which are denoted as follows:

$$\hat{a}_{\mathbf{k}\sigma} \equiv \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & \dots \\ 0 & 0 & \sqrt{2} & 0 & 0 & \dots \\ 0 & 0 & 0 & \sqrt{3} & 0 & \dots \\ 0 & 0 & 0 & 0 & 0 & \dots \\ \multicolumn{6}{c}{\dotfill} \end{pmatrix}$$

$$\hat{a}_{\mathbf{k}\sigma}^{+} \equiv \begin{pmatrix} 0 & 0 & 0 & 0 & \dots \\ 1 & 0 & 0 & 0 & \dots \\ 0 & \sqrt{2} & 0 & 0 & \dots \\ 0 & 0 & \sqrt{3} & 0 & \dots \\ \multicolumn{5}{c}{\dotfill} \end{pmatrix} \qquad (36.25)$$

In the matrix element defining the transition probability a row corresponds to the initial state of the system, a column to the final state (see (32.42)). Thus, the matrix elements in $\hat{a}_{\mathbf{k}\sigma}^{+}$ correspond to transitions in which the number of quanta increases by one, and $\hat{a}_{\mathbf{k}\sigma}$ corresponds to transitions in which the number of quanta decreases

by one. Accordingly, $\hat{a}_{k\sigma}^{+}$ is termed a *creation operator*, and $\hat{a}_{k\sigma}$ an *annihilation operator*.

If in some state there are $N_{k\sigma}$ quanta, the matrix element of the creation operator is proportional to $(N_{k\sigma} + 1)^{1/2}$, while that of the annihilation operator is proportional to $(N_{k\sigma})^{1/2}$. The transition probability involves the square of the matrix element, and the emission and absorption probabilities involve $N_{k\sigma} + 1$ and $N_{k\sigma}$.

Consequently, even in the absence of quanta of a given state in the electromagnetic field ($N_{k\sigma} = 0$) they may be emitted by the radiating system. This is known as *spontaneous emission*. The component proportional to $N_{k\sigma}$ in the emission probability describes *stimulated emission*. The limiting transition to the classical theory corresponds to $N_{k\sigma} \to \infty$, so that the spontaneous part of quantum emission becomes negligibly small.

Let us now determine the probability of the spontaneous emission of a quantum by a certain system of charges. If $\hat{\mathbf{p}}$ in (36.22) is interpreted as an operator, the question may arise as to how to write it with respect to a vector potential: according to (36.23), the latter depends upon position and, hence, is in the most general case noncommutative with $\hat{\mathbf{p}}$. But thanks to the transverse condition, it turns out that the order of $\hat{\mathbf{p}}$ and $\hat{A}$ is immaterial. Indeed, since div $\hat{\mathbf{A}} = 0$,

$$\hat{\mathbf{p}}\hat{\mathbf{A}}\psi = \left(\cdot \frac{h}{i} \operatorname{div} \hat{\mathbf{A}} \right) \psi + \hat{\mathbf{A}}\hat{\mathbf{p}}\psi = \hat{\mathbf{A}}\hat{\mathbf{p}}\psi$$

Thus, the matrix element of the transition of a radiating system of charges from a state with the wave function ψ_m, in which it emits a quantum $h\omega_k$ with the wave vector $\mathbf{k}$ and polarization σ, to a state with the wave function ψ_n, is

$$\mathcal{H}'_{n\,h\omega_{k},\,m\,0}| = -\frac{e}{m}\left(\frac{2\pi h}{\omega_{k}}\right)^{1/2}\int \psi_n^* \left(\mathbf{e}_k^\sigma \hat{\mathbf{p}}\right) e^{-i\mathbf{kr}}\,\psi_m' dV \quad (36.26)$$

In (36.26) to the matrix element of $\hat{a}_{k\sigma}^{+}$ there corresponds the element $(a_{k\sigma}^{+})_{10}$, equal to unity. The other factors in (36.26) are found from (36.22)-(36.24).

The transition matrix element (36.26) should be substituted into the general equation (32.42), assuming the energy of the emitted quantum to be equal to the energy difference of the radiating system:

$$h\omega_k = E_m - E_n \equiv h\omega_{mn} \quad (36.27)$$

Equation (32.42) also involves the "weight" of the final state of the system, that is, the number of states of the electromagnetic field per unit energy interval for the case of one quantum with the given wave vector $\mathbf{k}$ in the field. This quantity is easily found from

(28.23), which was developed for the number of oscillations of any nature in a certain volume V (which we put equal to unity). Further, we replace dp_x by $h \, dk_x$, etc., and in addition go over to spherical coordinates in the wave vector space. Then we obtain from Eq. (28.23)

$$dN_k = \frac{k^2 \, dk \, d\Omega}{(2\pi)^3} \tag{36.28}$$

Here, we must also replace k by ω/c and divide it by the energy differential $h \, d\omega$. Thus, we obtain the probability of the emission, in unit time, of a quantum of frequency $\omega = c \mid \mathbf{k} \mid$ and polarization σ in the direction of $\mathbf{k}$:

$$\frac{dW}{dt} = \frac{2\pi}{h} \mid \mathcal{H'}_{n \, h\omega_{\mathbf{k}}, \, m \, 0} \mid^2 \frac{\omega_{\mathbf{k}}^2 \, d\Omega}{(2\pi)^3 \, c^3 h}$$

$$= \left| \int \psi_n^* \, (\mathbf{e}_{\mathbf{k}}^\sigma \hat{\mathbf{p}}) \, e^{-i\mathbf{k}\mathbf{r}} \psi_m \, dV \right|^2 \frac{e^2 \omega_{\mathbf{k}} \, d\Omega}{2\pi h c^3} \tag{36.29}$$

The energy emitted in unit time is obtained from (36.29) by multiplying by $h\omega$.

The Absorption Coefficient. Let a plane-parallel flux of quanta be impinging on a certain system analogous to the one capable of quantum emission, so that in unit time one quantum crosses a unit area in a unit frequency interval. We must determine the probability of the absorption of a quantum from that beam in unit time.

This problem is similar to the determination of the scattering cross section; the difference is that in the latter case all the particles are assumed to have exactly the same energies, while quanta are not assumed in advance to be strictly monochromatic.[14] They are distributed over a certain spectral interval $\Delta \omega$ so that $h \, \Delta \omega \gg \Delta E$, where ΔE is the uncertainty in the energy of the absorbing system due to interaction with the electromagnetic field. In other words, the necessary relationship $E_m - E_n = h\omega$ is satisfied somewhere within the energy interval $h\Delta \omega$. Thus stated, the problem corresponds to real observations of absorption lines in a continuous spectrum.

The same method may be used to compute the absorption probability as was used to develop the emission probability formula (36.29). In the present case, however, the transition is not from a discrete spectrum to a continuous spectrum (emission) but the reverse. In

---

[14] In particle scattering the energy conservation law holds between two continuous spectrum states, initial and final. In the absorption of a quantum, the final state has a discrete spectrum. To satisfy the energy conservation law the spectrum of the initial states must be assumed continuous. Thanks to absorption they have a finite lifetime and cannot have an exact energy value.

absorption the spectrum of initial states is continuous, and, for example, the quanta may be assumed to be distributed over the frequencies in such a way that in the frequency interval $d\omega$ there are $I(\omega)\, d\omega$ quanta impinging on the absorbing system. The transition probability should be averaged over both the spectrum of the incident quanta and their direction and polarization. In all other respects the computations of the transition probability are similar to the development of Eq. (32.38). Therefore the absorption probability differs from the emission probability in the following factors:

(i) Instead of the "weight" factor $\omega^2/(2\pi c)^3$ the absorption probability involves the spectral density of the initial state $I(\omega)$, where $\hbar\omega = E_m - E_n$. Furthermore, averaging over the direction of the incident quanta yields the factor $(4\pi)^{-1}$ (that is, the result is divided by the magnitude of the total solid angle), and averaging over the polarization yields the factor $1/2$.

(ii) Since the absorption probability is calculated for an incident quantum flux equal to $1/(\text{cm}^2\text{s})$, the required expression must be divided by c (because we put $V = 1$).

(iii) In accordance with (36.25), the absorption probability involves the factor $N_{\mathbf{k}\sigma}$ instead of $N_{\mathbf{k}\sigma} + 1$ in the emission probability. Indices m and n in the matrix element are interchanged, but by virtue of the Hermiticity of $\hat{\mathbf{p}}$, the square of the modulus of the matrix element is the same.

The number of quanta, $N_{\mathbf{k}\sigma}$, is replaced according to the spectral formula by $I(\omega)\, d\omega$, after which $I(\omega)$ must be put equal to unity. Since the absorption coefficient is determined as the probability of absorption for the spectral density $I(\omega)$ equal to unity, we finally find that it differs from the emission coefficient in that it involves $1/(8\pi c)$ instead of the factor $\omega^2/(8\pi^3 c^3)$.

If instead of the emission probability we take the energy emitted in unit time for $N_{\mathbf{k}\sigma} = 0$, its ratio to the absorption coefficient when there is one quantum in the field ($N_{\mathbf{k}\sigma} = 1$) is equal to $\hbar\omega^3/(\pi^2 c^3)$. This expression was obtained by Einstein in 1916, using statistical methods.

The Dipole Approximation. The expression (36.29) is written for the most general case. Let us now apply it to the emission of visible light by an atomic system. The wavelength of visible light is approximately 0.5×10^{-4} cm, and the dimensions of an atom are about 10^{-8} cm, the order of the atomic unit. Hence, over a length of the order of atomic dimensions the phase of an electromagnetic wave remains almost unchanged. Therefore, for some point of the atom, for example, the nucleus (for $\mathbf{r} = \mathbf{r}_0$), the factor $e^{i\mathbf{k}\mathbf{r}}$ in (36.29) can be taken outside the integral. In the integrand there remains $e^{i\mathbf{k}(\mathbf{r}-\mathbf{r}_0)}$. But since the exponent is of the order of one ten-thousandth, the exponential can be replaced by unity if the integral does not vanish

in such a substitution. The latter possibility will be discussed later. But first we shall show that the integral does not vanish if we neglect the change in wave phase within the atom. Then we find

$$\frac{dW}{dt} = \mid (\mathbf{e}_{\mathbf{k}}^{\sigma} \cdot \mathbf{p}_{nm}) \mid^2 \frac{e^2 \omega_{\mathbf{k}} \, d\Omega}{2\pi hc^3 m^2} \qquad (36.30)$$

But the matrix element of momentum is related to the matrix element of position as follows:

$$p_{x_{nm}} = m\dot{x}_{nm} = im\omega_{nm} x_{nm}$$

(see (27.13)). If we also take into account that er is the dipole moment of the system, $\mathbf{d}$, we arrive at the formula

$$\frac{dW}{dt} = \mid (\mathbf{e}_{\mathbf{k}}^{\sigma} \cdot \mathbf{d}_{nm}) \mid^2 \frac{\omega_{\mathbf{k}}^3 \, d\Omega}{2\pi hc^3} \qquad (36.31)$$

By neglecting the change in wave phase over the dimensions of the radiating system we in effect neglected the retardation effects. Therefore, the radiation probability depends on the change in dipole moment in the same way as it does in the classical radiation theory (Sec. 20).

Suppose now that we have to find the probability of the emission of a quantum with any polarization, not the given polarization σ, but with a given direction $\mathbf{k}$. We draw a plane through vectors $\mathbf{k}$ and $\mathbf{d}_{nm}$ and assume that one polarization ($\sigma = 1$) corresponds to this plane, and another ($\sigma = 2$) corresponds to a plane perpendicular to it. Then $\mathbf{e}_{\mathbf{k}}^{(2)}\mathbf{d}_{nm} = 0$, and $\mathbf{e}_{\mathbf{k}}^{(1)}\mathbf{d}_{nm} = d_{nm} \sin \theta$, where θ is the angle between $\mathbf{k}$ and $\mathbf{d}_{nm}$.

To find the total energy radiated in unit time we must integrate the obtained expression over the solid angle Ω. Then after multiplying by $h\omega$ we get

$$\frac{dE}{dt} = \frac{\mid d_{nm} \mid^2}{c^3} \omega_{nm}^4 \int \frac{\sin^2 \theta \, d\Omega}{2\pi} = \frac{4}{3} \frac{\mid d_{nm} \mid^2 \omega_{nm}^4}{c^3} \qquad (36.32)$$

In comparison with the classical formula (20.28), the quantum equation involves an extra 2 (the factor 4/3 instead of 2/3). This is explained as follows. We represent the classical dipole moment varying according to the harmonic law as

$$\mathbf{d} = \mathbf{d}_1 e^{i\omega t} + \mathbf{d}_1^* e^{-i\omega t}, \qquad \ddot{\mathbf{d}} = -\omega^2 (\mathbf{d}_1 e^{i\omega t} + \mathbf{d}_1^* e^{-i\omega t})$$

The terms $\mathbf{d}_1 e^{i\omega t}$ and $\mathbf{d}_1^* e^{-i\omega t}$ depend upon time the same as the matrix elements d_{mn} and d_{nm}. We form the time average of $(\ddot{\mathbf{d}})^2$. The terms involving $e^{2i\omega t}$ and $e^{-2i\omega t}$ are eliminated in the averaging, leaving

$$\langle(\ddot{\mathbf{d}})^2\rangle = \omega^4 2\mathbf{d}_1\mathbf{d}_1^* = 2\omega^4 \mid d_1 \mid^2$$

The quantum formula corresponds to the mean radiation intensity $\hbar\omega$ (dW/dt), so that the 2 is associated with the time averaging of the square of the dipole moment.

Equation (36.32) confirms what was said in Section 21 about the stability of the atom following from quantum theory: radiation is always associated with an atom's transition from one state to another, but no states exist in which the energy would be less than the ground state, hence an atom may remain in the ground state indefinitely.

Selection Rules. Let us now determine the conditions at which the dipole moment matrix element does not vanish in a given transition. These conditions, imposed on the change of state of an atom in radiation, are known as *selection rules*.

We start with the simplest case of a one-electron transition in an atom. We furthermore assume that the atom is light and we need not take spin-orbit coupling into account. Then the spin function separates out of the total wave function as a factor. But since the matrix element is taken only with respect to the coordinate, the condition for the spin is that its projection does not change in the transition.

The principal quantum number imposes no restrictions: any variation is possible without the matrix element vanishing. There remain the orbital and magnetic quantum numbers.

In Exercise 4, Section 34, it was pointed out that the operator $\hat{\boldsymbol{\gamma}}$ satisfies the same commutation relations with the angular momentum operator as those which govern the position operator in a commutation with the particle's orbital angular momentum operator. But the condition that the matrix element $(\gamma)_{xyz}$ does not vanish was developed solely from the commutation relations It follows that the coordinate has the same nonzero matrix elements with respect to the angular momentum eigenvalues and projection as $\hat{\boldsymbol{\gamma}}$.

This leads to the following selection rules for the dipole moment components:

(i) The electron's orbital quantum number, l, can only vary in dipole radiation by ± 1; if it was l in the initial state it can be only $l + 1$ or $l - 1$ in the final state. In other words, $\Delta l = \pm 1$.

(ii) If the z axis is taken as the polar axis (the angular momentum quantization axis), then only the matrix elements $e\,(z)_{kk}$, which are diagonal with respect to the magnetic quantum number, are other than zero. For them $\Delta k = 0$.

The matrix elements $e(x \pm iy)_{k'k}$ are not zero when $k' - k \equiv$ $\equiv \Delta k = \pm 1$, according to the sign in $x \pm iy$. This rule means the following. If the radiation is polarized along the z axis, the electric field vector is determined by the zth component of the dipole moment. Hence, the selection rule $\Delta k = 0$ refers to such radiation. If

the radiation is circularly polarized in the x, y-plane, the rule for its circular components, the electric field of which is proportional to $e\,(x \pm iy)$, is $\Delta k = \pm 1$. For all three types of polarization (plane, and circular in both directions) the selection rule $\Delta l = \pm 1$ holds.

Let us now explain the physical meaning of the selection rules. Quantization of the electromagnetic field was achieved by a plane-wave expansion, when each separate harmonic wave was characterized by a definite momentum and polarization. If an expansion in spherical waves is effected in a spherical cavity with mirror walls, we find that each wave is characterized by angular momentum and parity in accordance with the general laws of isotropy of space. Here, the angular momentum is given by Eq. (15.28).

Spherical waves are represented quite independently of quantum theory with the help of the Y_L^K functions, where L and K are integers such that $-L \leqslant K \leqslant L$. This is the system of orthogonal functions on the surface of a unit sphere corresponding to rotation symmetry. The zth projection of the angular momentum is found to be

$$M_z = \frac{E}{\omega} K \qquad (36.33)$$

If we then quantize, for a separate quantum (photon) we obtain $E = h\omega$, so that the projection of the angular momentum of a photon, as of any particle, assumes only integral values. In addition to the angular momentum projection and energy, quantum mechanics deals with the angular momentum square M^2, the eigenvalues of which are $h^2 L\,(L+1)$, and $K \leqslant L$.

Electromagnetic waves are transverse and cannot have spherical symmetry. Accordingly, the quantum number L is never zero: only the function Y_0^0 is spherically symmetrical. The least possible value is $L = 1$. The symmetry and spatial configuration of the field correspond to electric or magnetic dipole radiation. For large values of L there are two multipole types for each L.

Electric and magnetic radiation differ in parity. Electrodipole radiation has parity -1, magnetodipole, $+1$, electroquadrupole, $+1$, magnetoquadrupole, -1, etc., alternately. This rule is easily explained as applied to dipole radiation. The electric dipole moment is a vector, the magnetic dipole moment is a pseudovector (see Sec. 15). In an inversion, vector components change their signs, while pseudovector components do not. This affects their parities. Each successive multipole order reverses the parity, insofar as it corresponds to the appearance of an additional factor **r** in the multipole moment.

In the case of a quantum, the angular momentum cannot be separated into orbital and spin angular momenta: this is possible only in nonrelativistic theory, while a quantum is a relativistic particle and its velocity is equal to c. But since its angular momentum cannot

be less than unity, we may say that its intrinsic angular momentum is equal to unity. However, the analogy with the angular momentum of a particle is not complete. For the case of $l = 1$, a particle has three states with the angular momentum projections k: -1, 0, and 1. A quantum with unity angular momentum is a dipole and can occur in only two states, electrodipole and magnetodipole.

Note the following. In a plane-wave expansion the state of a quantum is defined by four numbers, k_x, k_y, k_z, and σ. In a spherical-wave expansion we also obtain four numbers: ω, L, K, and parity. Like σ, parity takes on two values.

It is now not difficult to interpret the selection rules as a consequence of the angular momentum and parity conservation laws. The established rules refer to electrodipole radiation. Consequently, the angular momentum of a radiating electron cannot change by more than unity ($|\Delta l| \leqslant 1$). The parity must change, since the electrodipole quantum is odd. Hence, there can never be $\Delta l = 0$.

These selection rules are easily generalized for the case of sufficiently strong spin-orbit coupling and the electron state being characterized not separately by l and σ_z but by the total angular momentum j. Since angular momentum is not directly associated with parity, $\Delta j = 0$ is not prohibited in electric dipole radiation. Only a $(j = 0) \rightarrow (j = 0)$ transition is strictly prohibited, because it is incompatible with the angular momentum conservation law. The angular momentum of a quantum cannot be less than unity, therefore the $0 \rightarrow 0$ transition is prohibited in all approximations.

Let us consider many-electron atoms. If the spin-orbit coupling is not large the atom's spin function separates out as a factor of the spatial function. For example, in helium the total spin of the two electrons may be equal to unity, the ortho-state (Sec. 33), or to zero, the para-state. A transition from one to the other is prohibited because the wave functions of the ortho- and para-states are orthogonal, and the electric dipole moment does not depend on spin operators. Therefore, the dipole radiation spectrum of helium as it were separates into two spectra: one belonging wholly to orthohelium, the other to parahelium. This was known before quantum mechanics was developed, and it was explained by Heisenberg on the basis of wave function symmetry.

The selection rules for the total angular momentum J of an atom are similar to the selection rules for the angular momentum of a separate electron, j. The same is true of the selection rules for the zth projection of the total angular momentum.

Transitions prohibited for electrodipole quanta may be accompanied by emission of magnetodipole or higher multipole quanta. Corresponding to this in optical spectra is the lower line intensity according to the expansion of the wave amplitude in powers of the small quantities v/c and r/λ (see Sec. 20). For that reason excited

states of atoms or nuclei capable only of high multipole-order radiation transitions have long lifetimes and small energy level widths.

The Zeeman Multiplet. Let us apply the obtained selection rules to the Zeeman effect, that is, examine the spectrum of an atom in a magnetic field. We start with the case of a strong field, in which the splitting of spectral lines provides a simpler picture than in a weak field.

In accordance with Eq. (33.52), both multiplet levels between which the transitions take place are split. Assigning the spin orbital angular momentum projections involved in this equation all permitted values, we see that all the splitting components are equidistant and multiples of integers $L_z + 2S_z$, and because some values of $L_z + 2S_z$ are obtained in several ways, the levels corresponding to them are degenerate.

For example, if $L = 1$, $S = 1/2$, the following set of values of the sum obtains: $1 + 1 = 2$, $1 - 1 = 0$, $0 + 1 = 1$, $0 - 1 = -1$, $-1 + 1 = 0$, $-1 - 1 = -2$. Splitting occurs into a total of five levels, of which the zero level is two-fold degenerate, that is, six states occur. In a weak field, **L** and **S** first combine into **J**, which assumes two values: $J = 3/2$ and $J = 1/2$. Their multiplicities are 4 and 2—a total of 6 states, as in a strong field.

Suppose radiation is observed perpendicular to a magnetic field. The polarization vector of the radiation is perpendicular to the direction of propagation, that is, it is directed either along the magnetic field or in the third perpendicular direction, say along the x axis, if the magnetic field is parallel to the z axis. The selection rules differ for radiation polarized along z and along x. If spin-orbit coupling can be neglected, then S_z is separately conserved in any polarization. But in that case, for radiation polarized along z both S_z and L_z are conserved. Consequently, when a transition occurs between states split according to Eq. (33.52), these numbers cancel out, yielding one unsplit line.

A wave polarized along the x axis can be represented as a superposition of two circularly polarized waves of opposite sense. The selection rule for the corresponding lines is that L_z can vary only by ± 1. Consequently, radiation polarized along the x axis has a frequency differing from the initial one by $\pm e \mid \mathbf{H} \mid /(2mc)$. Thus, irrespective of the total number of splitting components of every level in a magnetic field, a spectral line in a strong magnetic field splits into three lines the separation of which is equal to $e \mid \mathbf{H} \mid /(2mc)$, that is, Larmour's frequency for the given field.

If we drill a hole in the armature of an electromagnet, it is possible to observe radiation along the field. It is circularly polarized in the x,y-plane. The selection rules for clockwise and counterclockwise polarization correspond to a variation of L_z by ± 1, so

that two lines are observed at a distance $e|\mathbf{H}|/(2mc)$ from the middle. When the magnetic field is switched on, the initial line splits into two lines, the separation between which is equal to the double Larmour frequency. The splitting pattern is exactly the same in the classical vibrational motion of a charge placed in a magnetic field. This problem was examined in Exercise 4, Section 20.

The splitting of spectral lines in a magnetic field was discovered by Zeeman before quantum mechanics appeared, which is why at the time the theoretical explanation of the effect corresponded to the classical problem, where it is assumed that the charge is in vibrational motion.

But such a spectral pattern is observed only in a strong magnetic field, which produces greater splitting than the separation between multiplet levels. In these conditions the Zeeman effect is said to be *normal*, because superficially it corresponds to the theoretical notions of the time when it was discovered. Note that a field that is strong for one multiplet may still be weak for another.

In a weak magnetic field the Zeeman effect is said to be *anomalous*. The spectral pattern is quite unlike the classical. First of all, the number of splitting components may differ from the normal. Their separations may also be quite different.

Let us consider, as an example, the anomalous Zeeman effect for the so-called D line of sodium vapour. In the absence of an external magnetic field the line is double. It corresponds to two transitions: $^2P_{1/2} \rightarrow {}^2S_{1/2}$ and $^2P_{3/2} \rightarrow {}^2S_{1/2}$. The ^2P level has orbital angular momentum unity and half-integral spin. Hence the total angular momentum J may assume two values: $J = 1 + 1/2 = 3/2$, and $J = 1 - 1/2 = 1/2$. This yields the doublet structure of the ^2P level in the absence of an external field. The $^2S_{1/2}$ level cannot split, since its orbital angular momentum is zero. A double D line appears in the sodium spectrum in a transition from a doublet to a singlet level. The separation for its frequency components is approximately one-thousandth of the mean frequency of the doublet. The $^2P_{3/2}$ level is higher than the $^2P_{1/2}$ level.

Let us calculate the Lande factors for the three levels. From (33.51) we have:

(1) $^2P_{3/2}$: $J = 3/2$, $L = 1$, $S = 1/2$

$$1 + \frac{1}{2}\, \frac{3/2 \times 5/2 + 1/2 \times 3/2 - 1 \times 2}{3/2 \times 5/2} = \frac{4}{3}$$

(2) $^2P_{1/2}$: $J = 1/2$, $L = 1$, $S = 1/2$

$$1 + \frac{1}{2}\, \frac{1/2 \times 3/2 + 1/2 \times 3/2 - 1 \times 2}{1/2 \times 3/2} = \frac{2}{3}$$

(3) $^2S_{1/2}$: $J = 1/2$, $L = 0$, $S = 1/2$

$$1 + \frac{1}{2}\frac{1/2 \times 3/2 + 1/2 \times 3/2}{1/2 \times 3/2} = 2$$

We shall measure the splitting energy E' in units of $\beta|\mathbf{H}|$ so that

$$E' = gJ_z \tag{36.34}$$

where g is the corresponding Lande factor. We then obtain the following splitting pattern for the levels $^2P_{3/2}$, $^2P_{1/2}$, and $^2S_{1/2}$:

(1) $E'\left(-\dfrac{3}{2}\right) = -2$, $E'\left(-\dfrac{1}{2}\right) = -\dfrac{2}{3}$,

       $E'\left(\dfrac{1}{2}\right) = \dfrac{2}{3}$, $E'\left(\dfrac{3}{2}\right) = 2$

(2) $E'\left(-\dfrac{1}{2}\right) = -\dfrac{1}{3}$, $E'\left(\dfrac{1}{2}\right) = \dfrac{1}{3}$

(3) $E'\left(-\dfrac{1}{2}\right) = -1$, $E'\left(\dfrac{1}{2}\right) = 1$

Let us find what the spectrum looks like. We start with the transition $^2P_{1/2} \to {}^2S_{1/2}$. Vibrations polarized along the field are subject to the selection rule $\Delta J_z = 0$. Hence, their frequencies are shifted with respect to the median position by

$$E'\left({}^2P_{1/2},\ -\frac{1}{2}\right) - E'\left({}^2S_{1/2},\ -\frac{1}{2}\right) = -\frac{1}{3} + 1 = \frac{2}{3}$$

$$E'\left({}^2P_{1/2},\ \frac{1}{2}\right) - E'\left({}^2S_{1/2},\ \frac{1}{2}\right) = \frac{1}{3} - 1 = -\frac{2}{3}$$

Unlike the normal Zeeman effect, we have also obtained a double line for radiation polarized along the field.

For polarizations perpendicular to the field we have

$$E'\left({}^2P_{1/2},\ -\frac{1}{2}\right) - E'\left({}^2S_{1/2},\ \frac{1}{2}\right)$$

$$= -\frac{1}{3} - 1 = -\frac{4}{3} \quad \text{(clockwise polarization)}$$

$$E'\left({}^2P_{1/2},\ \frac{1}{2}\right) - E'\left({}^2S_{1/2},\ -\frac{1}{2}\right)$$

$$= \frac{1}{3} + 1 = \frac{4}{3} \quad \text{(counterclockwise polarization)}$$

Refferring now to the transition $^2P_{3/2} \to {}^2S_{1/2}$, if the vibration is polarized along the field, we have, of course, two lines again, but with a different separation from the median position:

$$E'\left({}^2P_{3/2},\ -\frac{1}{2}\right) - E'\left({}^2S_{1/2},\ -\frac{1}{2}\right) = -\frac{2}{3} + 1 = \frac{1}{3}$$

$$E'\left({}^2P_{3/2},\ \frac{1}{2}\right) - E'\left({}^2S_{1/2},\ \frac{1}{2}\right) = \frac{2}{3} - 1 = -\frac{1}{3}$$

For clockwise polarization we obtain

$$E'\left({}^2P_{3/2},\ -\frac{3}{2}\right) - E'\left({}^2S_{1/2},\ -\frac{1}{2}\right) = -2 + 1 = -1$$

$$E'\left({}^2P_{3/2},\ -\frac{1}{2}\right) - E'\left({}^2S_{1/2},\ \frac{1}{2}\right) = -\frac{2}{3} - 1 = -\frac{5}{3}$$

Correspondingly, for counterclockwise polarization the splitting is characterized by the numbers 1 and 5/3.
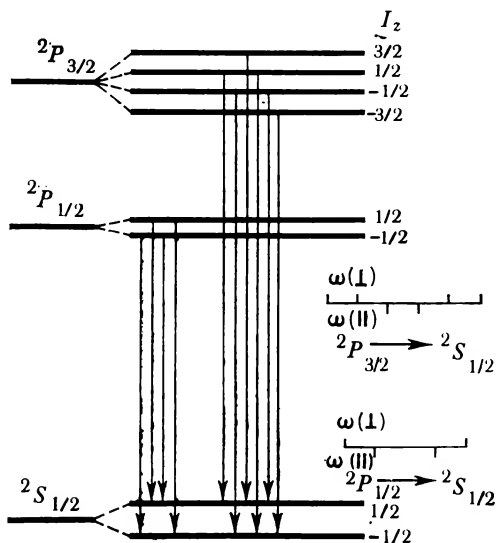
Figure 41

Thus, one component of the D line splits into six Zeeman components, and the other into four. The Zeeman effect remains anomalous as long as the magnetic field is weaker than 5000 statoersted. The splitting pattern is presented in Figure 41.

EXERCISES

1. Develop the splitting pattern for the multiplet ${}^3P$ and ${}^3S$ and transitions in a weak and a strong magnetic field.

2. Calculate the linear momentum of an electromagnetic field in vacuum with the help of the normal field coordinates $Q_{\mathbf{k}}^{\sigma}$.

Solution. Substitute the electric and magnetic fields according to (36.6) and (36.7) into the expression (15.25):

$$\mathbf{p} = \frac{1}{4\pi c} \int (\mathbf{E} \times \mathbf{H}) \, dV$$

After integrating over the volume we obtain, making use of (36.11),

$$\mathbf{p} = \frac{1}{4\pi c^2} \sum_{\mathbf{k}, \sigma, \sigma'} \omega_{\mathbf{k}} \{ [A_{\mathbf{k}}^{\sigma} \times (\mathbf{k} \times A_{-\mathbf{k}}^{\sigma'})] + [A_{\mathbf{k}}^{\sigma} \times (\mathbf{k} \times A_{\mathbf{k}}^{\sigma'*})]$$
$$+ [A_{\mathbf{k}}^{\sigma*} \times (\mathbf{k} \times A_{\mathbf{k}}^{\sigma'})] + [A_{\mathbf{k}}^{\sigma*} \times (\mathbf{k} \times A_{-\mathbf{k}}^{\sigma'*})] \}$$

We transform the double vector products and get

$$\mathbf{p} = \frac{1}{4\pi c^2} \sum_{\mathbf{k}, \sigma} \omega_{\mathbf{k}} [\mathbf{k} (A_{\mathbf{k}}^{\sigma} \cdot A_{-\mathbf{k}}^{\sigma}) + \mathbf{k} (A_{\mathbf{k}}^{\sigma*} \cdot A_{-\mathbf{k}}^{\sigma*}) + 2\mathbf{k} (A_{\mathbf{k}}^{\sigma*} \cdot A_{\mathbf{k}}^{\sigma})]$$

The quantities $\mathbf{k} (A_{\mathbf{k}}^{\sigma} \cdot A_{-\mathbf{k}}^{\sigma})$ and $\mathbf{k} (A_{\mathbf{k}}^{\sigma*} \cdot A_{-\mathbf{k}}^{\sigma*})$ are odd functions of $\mathbf{k}$, and after summation over all $\mathbf{k}$ they vanish, leaving only

$$\mathbf{p} = \frac{1}{2\pi c^2} \sum_{\mathbf{k}, \sigma} \omega_{\mathbf{k}} \mathbf{k} (A_{\mathbf{k}}^{\sigma*} \cdot A_{\mathbf{k}}^{\sigma})$$

Substitution of the normal field coordinates from (36.17a) and (36.17b) yields

$$\mathbf{p} = \sum_{\mathbf{k}, \sigma} \frac{\mathbf{k}}{\omega_{\mathbf{k}}} \frac{1}{2} [(P_{\mathbf{k}}^{\sigma})^2 + \omega_{\mathbf{k}}^2 (Q_{\mathbf{k}}^{\sigma})^2] = \sum_{\mathbf{k}, \sigma} \hbar \mathbf{k} \left(N_{\mathbf{k}\sigma} + \frac{1}{2} \right)$$

so that (14.13) gives the relationship between the momentum and energy of a quantum.

3. Obtain the commutation relations for the operators $\hat{a}_{\mathbf{k}\sigma}$ and $\hat{a}_{\mathbf{k}\sigma}^+$, and determine $\hat{a}_{\mathbf{k}\sigma}^+ \hat{a}_{\mathbf{k}\sigma}$ and $\hat{a}_{\mathbf{k}\sigma} \hat{a}_{\mathbf{k}\sigma}^+$.

Answer. $\hat{a}_{\mathbf{k}\sigma}^+ \hat{a}_{\mathbf{k}\sigma}$ and $\hat{a}_{\mathbf{k}\sigma} \hat{a}_{\mathbf{k}\sigma}^+$ are diagonal; $\hat{a}_{\mathbf{k}\sigma}^+ \hat{a}_{\mathbf{k}\sigma} = N_{\mathbf{k}\sigma} + 1$ and $\hat{a}_{\mathbf{k}\sigma} \hat{a}_{\mathbf{k}\sigma}^+ = N_{\mathbf{k}\sigma}$. From this we have the commutation relation

$$\hat{a}_{\mathbf{k}\sigma}^+ \hat{a}_{\mathbf{k}\sigma} - \hat{a}_{\mathbf{k}\sigma} \hat{a}_{\mathbf{k}\sigma}^+ = 1$$

For different $\mathbf{k}$, σ and $\mathbf{k}'$, σ' the operators $\hat{a}_{\mathbf{k}\sigma}^+$ and $\hat{a}_{\mathbf{k}'\sigma'}$ commute. The operators $\hat{a}_{\mathbf{k}\sigma}$ and $\hat{a}_{\mathbf{k}'\sigma'}$ apparently always commute, the same as $\hat{a}_{\mathbf{k}\sigma}^+$ and $\hat{a}_{\mathbf{k}'\sigma'}^+$.

37

THE DIRAC EQUATION

Spinors in Four Dimensions. In the preceding section we developed the quantum theory of a relativistic particle, the light quantum. We were not concerned with the relativistic invariance of the equations, because it was inherent in the classical system, in Maxwell's

equations. There is no such initial wave equation for the electron, because there is nothing in the limiting classical case that corresponds to its wave function (Sec. 36). Here we must from the outset require that the equations be relativistically invariant.

In any case, it is obvious that the electron wave equation, which is invariant with respect to the Lorentz transformations, must of necessity take account of spin, because spin-orbit interaction is a relativistic effect. As was shown in Section 30, the wave function of a particle with half-integral spin is a two-component spinor quantity, which in rotations of the coordinate axes transforms through half-angles. Since the Lorentz transformation can be likened to a rotation of the coordinate system (Sec. 13), the definition of a spinor should be extended to rotations in four-dimensional space, by analogy with the extension of the vector definition to four-vectors.

In the most general case the transformation of a spinor in the rotations of a coordinate system takes place according to the law

$$\xi_1' = \alpha\xi_1 + \beta\xi_2$$
$$\xi_2' = \gamma\xi_1 + \delta\xi_2 \tag{37.1}$$

Then, as is readily apparent, the following combination of the components of two different spinors, ξ and η,

$$\xi_1'\eta_2' - \xi_2'\eta_1' = \xi_1\eta_2 - \xi_2\eta_1 \tag{37.2}$$

remains invariant if

$$\alpha\delta - \beta\gamma = 1 \tag{37.3}$$

If ξ_1 and ξ_2 are the components of a spinor wave function, then $|\xi_1|^2 + |\xi_2|^2$ is the probability density of the particles occurring at the given point in space. In three dimensions this is a scalar quantity, which imposed certain restrictions on the transformation coefficients, namely $\alpha^* = \delta$ and $\beta^* = -\gamma$.

In four dimensions the probability density should be considered as the fourth component of a vector, therefore, if in the ordinary rotations of the coordinate axes we also take into account the Lorentz rotations, then the transformations of a certain spinor (ξ_1, ξ_2) and the transformations of the complex conjugate spinor are not linked by the same relationships as in three-dimensional space. Accordingly, in future we shall denote such spinors by an asterisk (another convention is to place a dot over the spinor index).

Suppose a pure Lorentz rotation is performed through an imaginary angle ω the tangent of which is equal to iV/c. Then, as can be seen from (30.42), ξ_1 and ξ_2 are replaced by $\xi_1 e^{i\omega/2}$ and $\xi_2 e^{-i\omega/2}$. But since ω is an imaginary angle, the rotation equations for the spinor components take the form

$$\xi_1' = \xi_1 e^{-|\omega|/2}, \qquad \xi_2' = \xi_2 e^{|\omega|/2} \tag{37.4}$$

So as to arrive sooner at the standard notation of Dirac's theory we shall assume that the relative velocity V of the reference frames is directed along the z axis. Then the Lorentz transformations (13.17) and (13.18) can be expressed as

$$z' = \frac{z - Vt}{(1 - V^2/c^2)^{1/2}}, \qquad t' = \frac{t - Vx/c^2}{(1 - V^2c^2)^{1/2}}$$

We form linear combinations

$$ct' \pm z' = (ct \pm z)\, \frac{1 \pm V/c}{(1 - V^2/c^2)^{1/2}}$$

Substituting $V/c = -i \tan \omega$ into this, we find

$$ct' \pm z' = (ct \pm z)\, e^{\mp i\omega} = (ct \pm z)\, e^{\pm |\omega|} \qquad (37.5)$$

since ω is a purely imaginary quantity.

Since in the present case ξ_1^* transforms exactly like ξ_1 (because $e^{-|\omega|/2}$ is real), we arrive at the following correspondence between the transformations of the products of the spinor components multiplied by their complex conjugates, and the vector components:

$$\xi_1'^* \xi_1' = \xi_1^* \xi_1 e^{-|\omega|}, \qquad ct' + z' = (ct + z)\, e^{|\omega|}$$

$$\xi_2'^* \xi_2' = \xi_2^* \xi_2 e^{|\omega|}, \qquad ct' - z' = (ct - z)\, e^{-|\omega|}$$

This Lorentz transformation does not affect the x and y coordinates and does not alter the products $\xi_2^* \xi_1$ and $\xi_1^* \xi_2$. To establish the correspondence between them, consider a spatial rotation about the z axis:

$$x' = x \cos \varphi + y \sin \varphi$$

$$y' = -x \sin \varphi + y \cos \varphi$$

Multiply the second equation by $\pm i$ and add with the first to get

$$x' \pm iy' = (x \pm iy)\, e^{\mp i\varphi} \qquad (37.6)$$

From this we arrive at the correspondence

$$\xi_2^{*\prime} \xi_1' = \xi_2^* \xi_1 e^{i\varphi}, \qquad x' + iy' = (x + iy)\, e^{-i\varphi}$$

$$\xi_1^{*\prime} \xi_2' = \xi_1^* \xi_2 e^{-i\varphi}, \qquad x' - iy' = (x - iy)\, e^{i\varphi}$$

Thus, if there is given a four-vector momentum $p_t,\ p_x,\ p_y,\ p_z$ ($p_t \equiv E/c$) and a spinor $(\xi_1,\ \xi_2)$, the following relativistically invariant quantity is developed from them:

$$\xi_1^* (p_t + p_z)\, \xi_1 + \xi_2^* (p_t - p_z)\, \xi_2$$
$$+ \xi_2^* (p_x + ip_y)\, \xi_1 + \xi_1^* (p_x - ip_y)\, \xi_2 \qquad (37.7)$$

Here we have written the components of the four-vector of momentum between the spinor components, since subsequently **p** should be considered as an operator.

The relativistically invariant Lagrangian, from which follows the electron wave equation, must be expressed in terms of the relativistically invariant quantity (37.7). But the Lagrangian apparently involves, in addition to the momentum of the electron, also its mass, which is itself relativistically invariant and should therefore be multiplied by a like quantity.

As we saw at the beginning of this section, a scalar may be developed only from the components of two spinors, say ξ and η, and also from the conjugates ξ^*, η^*. These scalars are $\xi_1\eta_2 - \xi_2\eta_1$ and $\xi_1^*\eta_2^* - \xi_2^*\eta_1^*$. Then the Lagrangian also involves an expression of the form (37.7) constructed from the components of η.

The labelling of the components with asterisks is used to stress their complex conjugate transformation law with respect to (37.1). To go over to conventional quantum mechanical notation it is more convenient to put

$$\xi_1 = \psi_1, \qquad \xi_1^* = \psi_1^*, \qquad \xi_2 = \psi_2, \qquad \xi_2^* = \psi_2^*$$
$$\eta_1^* = \psi_4, \qquad \eta_1 = \psi_4^*, \qquad \eta_2^* = \psi_3, \qquad \eta_2 = \psi_3^*$$

It should be remembered here that ψ_3 and ψ_4 transform not like ψ_1 and ψ_2, but according to complex conjugate equations.

Let us now write the Lagrangian, denoting the wave function in terms of $\psi_1, \ldots, \psi_4$:

$$L = \psi_1^*\,(p_t + \hat{p}_z)\,\psi_1 + \psi_2^*\,(p_t - \hat{p}_z)\,\psi_2 + \psi_2^*\,(\hat{p}_x + i\hat{p}_y)\,\psi_1$$
$$+ \psi_1^*\,(\hat{p}_x - i\hat{p}_y)\,\psi_2 + \psi_4^*\,(p_t + \hat{p}_z)\,\psi_4 + \psi_3^*\,(p_t - \hat{p}_z)\,\psi_3$$
$$+ \psi_4^*\,(\hat{p}_x + i\hat{p}_y)\,\psi_3 + \psi_3^*\,(\hat{p}_x - i\hat{p}_y)\,\psi_4$$
$$- mc\,(\psi_1\psi_3^* - \psi_2\psi_4^*) - mc\,(\psi_1^*\psi_3 - \psi_2^*\psi_4) \quad (37.8)$$

Note that the coefficients m can always be made the same by a due choice of factors multiplying the ψ components. Therefore in this sense (37.8) does not involve any restriction of generality.

The Dirac Equation. We now vary L with respect to ψ_1^*, ψ_2^*, ψ_3^*, and ψ_4^* and equate the variations to zero. We write the operators operating on these functions on the left, making use of their Hermiticity:

$$(p_t + \hat{p}_z)\,\psi_1 + (\hat{p}_x - i\hat{p}_y)\,\psi_2 - mc\psi_3 = 0$$
$$(p_t - \hat{p}_z)\,\psi_2 + (\hat{p}_x + i\hat{p}_y)\,\psi_1 + mc\psi_4 = 0$$
$$(p_t - \hat{p}_z)\,\psi_3 + (\hat{p}_x - i\hat{p}_y)\,\psi_4 - mc\psi_1 = 0$$
$$(p_t + \hat{p}_z)\,\psi_4 + (\hat{p}_x + i\hat{p}_y)\,\psi_3 + mc\psi_2 = 0 \quad (37.9)$$

This is the required *Dirac equation*. The expanded form of Eq. (37.9) is rarely used, symbolic matrix notation usually being employed. For that the Pauli matrices (30.31) are used. Let us suppose that these matrices operate similarly on each pair ψ_1, ψ_2, and ψ_3, ψ_4, so that from (30.32) we obtain

$$\hat{\sigma}_x \begin{pmatrix} \psi_1 \\ \psi_2 \end{pmatrix} = \begin{pmatrix} \psi_2 \\ \psi_1 \end{pmatrix}, \quad \hat{\sigma}_y \begin{pmatrix} \psi_1 \\ \psi_2 \end{pmatrix} = i \begin{pmatrix} -\psi_2 \\ \psi_1 \end{pmatrix},$$

$$\hat{\sigma}_z \begin{pmatrix} \psi_1 \\ \psi_2 \end{pmatrix} = \begin{pmatrix} \psi_1 \\ -\psi_2 \end{pmatrix}$$

$$\hat{\sigma}_x \begin{pmatrix} \psi_3 \\ \psi_4 \end{pmatrix} = \begin{pmatrix} \psi_4 \\ \psi_3 \end{pmatrix}, \quad \hat{\sigma}_y \begin{pmatrix} \psi_3 \\ \psi_4 \end{pmatrix} = i \begin{pmatrix} -\psi_4 \\ \psi_3 \end{pmatrix},$$

$$\hat{\sigma}_z \begin{pmatrix} \psi_3 \\ \psi_4 \end{pmatrix} = \begin{pmatrix} \psi_3 \\ -\psi_4 \end{pmatrix}$$

As can be seen from (37.9), we also need matrices that interchange the pairs ψ_1, ψ_2 and ψ_3, ψ_4. We denote these matrices $\hat{\rho}_1$, $\hat{\rho}_2$, and $\hat{\rho}_3$. They operate in exactly the same way as $\hat{\sigma}_x$, $\hat{\sigma}_y$, and $\hat{\sigma}_z$, but with respect to each pair, as though we were using the notation $\psi_1 = \psi_1^1$, $\psi_2 = \psi_2^1$, $\psi_3 = \psi_1^2$, $\psi_4 = \psi_2^2$. In other words, they would operate only on the superscript. The commutativity of $\hat{\rho}$ and $\hat{\sigma}$ is readily apparent from this, but we shall not use the two-index notation. It has been cited only to show the commutativity of $\hat{\rho}$ and $\hat{\sigma}$ more vividly.

Now consider the system (37.9). The components of ψ attached to p_t are in normal sequence, hence we have a unit matrix (which we do not write). The components of ψ attached to $\hat{p}_z$ also appear in normal sequence, but with different signs. Here we must write $\hat{\rho}_3\hat{\sigma}_z$, then ψ_2 and ψ_3 acquire minus signs. Attached to $\hat{p}_x$ is the matrix $\hat{\sigma}_x$, which does not interchange the first and second component pairs, and $\hat{\sigma}_y$ is attached to $\hat{p}_y$ for the same reason. Finally, attached to m is the matrix $-\hat{\rho}_1\hat{\sigma}_z$. We also introduce the following abbreviated notation:

$$\hat{\alpha}_x = -\hat{\sigma}_x, \quad \hat{\alpha}_y = -\hat{\sigma}_y, \quad \hat{\alpha}_z = -\hat{\rho}_3\hat{\sigma}_z, \quad \hat{\beta} = \hat{\rho}_1\hat{\sigma}_z \quad (37.10a)$$

Then in the abbreviated form the system (37.9) is written as follows:

$$p_t\psi = (\hat{\alpha}_x\hat{p}_x + \hat{\alpha}_y\hat{p}_y + \hat{\alpha}_z\hat{p}_z + \hat{\beta}mc)\,\psi = \hat{\boldsymbol{\alpha}}\hat{\mathbf{p}}\psi + \hat{\beta}mc\psi \quad (37.11)$$

The operators $\hat{\sigma}_x, \hat{\sigma}_y$, and $\hat{\sigma}_z$, as well as the operators $\hat{\rho}_1, \hat{\rho}_2$, and $\hat{\rho}_3$, possess the following properties (see Sec. 30):

$$\hat{\sigma}_x^2 = \hat{\sigma}_y^2 = \hat{\sigma}_z^2 = 1$$

$$\hat{\sigma}_x\hat{\sigma}_y + \hat{\sigma}_y\hat{\sigma}_x = \hat{\sigma}_x\hat{\sigma}_z + \hat{\sigma}_z\hat{\sigma}_x = \hat{\sigma}_y\hat{\sigma}_z + \hat{\sigma}_z\hat{\sigma}_y = 0$$

$$\hat{\rho}_1^2 = \hat{\rho}_2^2 = \hat{\rho}_3^2 = 1$$

$$\hat{\rho}_1\hat{\rho}_2 + \hat{\rho}_2\hat{\rho}_1 = \hat{\rho}_1\hat{\rho}_3 + \hat{\rho}_3\hat{\rho}_1 = \hat{\rho}_2\hat{\rho}_3 + \hat{\rho}_3\hat{\rho}_2 = 0$$

Using this, we easily find the analogous properties of $\hat{\alpha}$ and $\hat{\beta}$:

$$\hat{\alpha}_x^2 = \hat{\alpha}_y^2 = \hat{\alpha}_z^2 = \hat{\beta}^2 = 1 \tag{37.12}$$

$$\hat{\alpha}_x\hat{\alpha}_y + \hat{\alpha}_y\hat{\alpha}_x = \hat{\alpha}_x\hat{\alpha}_z + \hat{\alpha}_z\hat{\alpha}_x = \hat{\alpha}_y\hat{\alpha}_z + \hat{\alpha}_z\hat{\alpha}_y$$
$$= \hat{\alpha}_x\hat{\beta} + \hat{\beta}\hat{\alpha}_x = \hat{\alpha}_y\hat{\beta} + \hat{\beta}\hat{\alpha}_y = \hat{\alpha}_z\hat{\beta} + \hat{\beta}\hat{\alpha}_z = 0 \tag{37.13}$$

We apply the operator $p_t = -(h/ic)(\partial/\partial t)$ to the left-hand side of (37.11), and the operator $\hat{\alpha}\mathbf{p} + \hat{\beta}mc$ to the right-hand side.[15] Then, using (37.12) and (37.13), we obtain

$$p_t^2\psi = \hat{p}^2\psi + m^2c^2\psi \tag{37.14}$$

since all the products of the different operators α_i, β cancel out, and the squares of operators are equal to unity. Thus, all the operators interchanging the functions have disappeared, and we are left with a wave-type differential equation

$$\frac{1}{c^2}\frac{\partial^2\psi}{\partial t^2} - \nabla^2\psi = -\frac{m^2c^2}{h^2}\psi \tag{37.15}$$

Its relativistic invariance is obvious. Applied to a free particle, its solution must be sought in the form of a plane wave

$$\psi = \psi_0 e^{-iEt/h + i\mathbf{pr}/h} \tag{37.16}$$

whence follows the correct relativistic relationship between energy and momentum:

$$E^2 = c^2p^2 + m^2c^4 \tag{37.17}$$

Spin has also been eliminated from the wave equation (37.15), which refers to a one-component wave function. For this reason Eq. (37.15), which was independently enunciated by Schrödinger, Fock, Klein and Gordon as an obvious (at first glance) generalization of the nonrelativistic wave equation, does not apply to the electron without taking Dirac's general equation (37.11) into account.

---

[15] We do not write the "cap" (^) over p_t so as to stress that this quantity is not a usual quantum mechanical operator. The operator notation is appropriate if it is understood as the Hamiltonian divided by c.

We shall now show that the Dirac equation is invariant not only with respect to rotations and Lorentz transformations, as having been developed from the relativistically invariant Lagrangian, but also with respect to an inversion of the coordinate system. In an inversion, $\mathbf{p}$ transforms to $-\mathbf{p}$, so that the Dirac equation acquires the form

$$p_t\psi = -\hat{\boldsymbol{\alpha}}\,\hat{\mathbf{p}}\psi + mc\hat{\beta}\psi$$

We multiply both sides by $\hat{\beta}$ and take advantage of the fact that, from (37.13), $\hat{\beta}\hat{\alpha} = -\hat{\alpha}\hat{\beta}$. Then

$$p_t\,(\hat{\beta}\psi) = \hat{\boldsymbol{\alpha}}\,\hat{\mathbf{p}}\,(\hat{\beta}\psi) + mc\hat{\beta}\,(\hat{\beta}\psi) \tag{37.18}$$

Thus, the function $\hat{\beta}\psi$ satisfies the same equation as the initial function, ψ. But since $\hat{\beta}\psi$ is in principle equivalent to ψ, we may assert that the Dirac equation has been reduced to its initial form. Multiplication by $\hat{\beta}$ is a certain special unitary transformation of the wave function.

The known unitary transformation is usually performed in a way such that matrices $\hat{\alpha}$ and $\hat{\beta}$, while retaining their basic properties (37.12) and (37.13), are differently expressed in terms of matrices $\hat{\rho}$ and $\hat{\sigma}$. We shall not perform the transformation itself and simply write the new expressions similar to (37.10a):

$$\hat{\alpha}_x = \hat{\rho}_1\hat{\sigma}_x, \qquad \hat{\alpha}_y = \hat{\rho}_1\hat{\sigma}_y, \qquad \alpha_z = \hat{\rho}_1\hat{\sigma}_z, \qquad \hat{\beta} = \hat{\rho}_3 \tag{37.10b}$$

The Dirac equation employing such $\hat{\alpha}$ and $\hat{\beta}$ is more convenient for the limiting transition to the nonrelativistic wave equation. It is easy to verify that $\hat{\alpha}$ and $\hat{\beta}$ determined with the help of (37.10b), really do possess the properties (37.12) and (37.13), which alone is necessary to obtain the correct energy-momentum relation (37.17).

Energy Eigenvalues. It follows from Eq. (37.17) that

$$E = \pm\,(c^2p^2 + m^2c^4)^{1/2} \tag{37.19}$$

The energy eigenvalue of an electron determined from the Dirac equation may be negative as well as positive. In classical mechanics only the plus sign is taken, because free electrons do not have negative energies.

The absolute value of the square root in (37.19) is not less than mc^2, so that there is an energy gap of width $2mc^2$ to which the energy of the electron cannot belong. In classical equations all quantities vary continuously, therefore energy, once defined by a positive sign, does not cross the prohibited $2mc^2$ gap and always retains the appropriate sign. In other words, energy positively defined by initial conditions remains positive according to the equations of motion.

But in quantum theory discontinuous transitions between different states are also possible. For example, an electron with energy greater than mc^2 could emit a light quantum and remain with an energy smaller than $-mc^2$. However, electrons with negative energy and mass are not observed in nature. Their properties would be very strange indeed: while radiating they would dissipate their energy, collapsing, so to say, into a state with energy $E = -\infty$. Very soon all the electrons in the universe would have collapsed into that state, contrary to what we observe all around us.

Thus, the Dirac equation allows for the possibility of states which, on the one hand, cannot be rejected, because electrons could transfer to them from known states, while, on the other hand, electrons with negative energy nonetheless do not exist. At the same time the Dirac equation describes a variety of electron properties absolutely correctly: as we shall soon see, it yields a relationship between spin and magnetic moment that agrees with experiment, leads to the precise formula of the fine structure of the hydrogen atom, etc. Furthermore, mathematical investigations show that there is no substantially other relativistically invariant equation for a particle of half-integral spin and nonzero mass. Our derivation of the Dirac equation from the invariant expression for the Lagrangian developed from spinors points to this convincingly enough. It would therefore be wrong to reject the Dirac equation out of hand: it is much better to supplement it with some suitable physical hypothesis.

Dirac suggested reformulating the concept of vacuum. Formerly by vacuum was meant a state of the electromagnetic field without electric charges. Dirac felt it possible to describe as vacuum a state in which all levels with negative values are occupied by electrons. That this redefinition is not just an exercise in semantics and it carries physical meaning will be made apparent shortly.

If all levels with negative energy are occupied, then by Pauli's exclusion principle electrons cannot transfer to them from positive-energy states. Thus, Pauli's exclusion principle is essential for the relativistic theory to be able to describe electron properties at all. Therein is the substantiation of Pauli's principle as a necessary element of quantum mechanics. In nonrelativistic theory Pauli's principle exists simply as a supplementary postulate of the many-body problem.

In Section 36 we defined vacuum as a state of an electromagnetic field in which there are no quanta, in other words, the ground state with the least possible energy. In the same way, if all negative-energy levels are occupied, the remaining electrons can no longer reduce their energy by passing into negative-energy states, and consequently, a state with only occupied negative levels possesses the least possible energy with respect to the electrons. Such a state is termed the *electron field vacuum*.

Pair Production. The expression "electron field" is used by analogy with electromagnetic field for the following reason. The Dirac equation is essentially never applied to one electron: it always implies the existence of a "background", that is, negative-energy states occupied by other electrons. Otherwise that electron would itself pass into a negative-energy state.

But the "background" not only guarantees the electron from "falling"; its existence is manifested in a real physical process. In an external electromagnetic field, for example, close to a nucleus, a quantum with energy greater than $2mc^2$ is capable of catapulting an electron from a negative-energy state to a positive-energy state. The external field is necessary to satisfy the momentum conservation law. For the proof of this simple assertion see Exercise 1.

But the removal of an electron from the negative-energy state leaves a "hole", that is, an unoccupied level (cf. Sec. 33). In an electric field, electrons of negative mass (mass is of the same sign as energy) move not against the field, towards the anode, but along the field, towards the cathode. The "hole" moves together with them, thus behaving like a positively charged electron with positive mass.

Experimentally, the ejection of an electron from the negative-energy state should produce two charges, one negative and one positive. This theoretical prediction was subsequently confirmed by Anderson.

A positron and an electron may mutually annihilate on impact if the electron transfers to an unoccupied level in the negative-energy states. Its energy is transferred to the electromagnetic radiation in the form of two or three quanta. Annihilation in vacuum cannot produce one quantum, as this would violate the momentum conservation law, just as one quantum cannot produce an electron-positron pair in vacuum. But in a nuclear field one-quantum annihilation is possible.

Electrons and positrons are known as particles and antiparticles because of their ability for mutual annihilation. Antiprotons and antineutrons have also been discovered.

Owing to the "background", the relativistic quantum theory of the electron is in effect a theory not of a separate particle but of a field in which the number of particles is not defined. Depending on the energy, one or several pairs may appear in the field in addition to the one electron, much as quanta are emitted in an electromagnetic field. Only the total charge is strictly conserved, but not the number of particles.

If the energy is insufficient for the actual production of pairs, transitions may exist in which pairs appear and then vanish. The intermediate states are of such short duration that their energy is quite indeterminate, as in the case of an alpha-particle below the potential barrier on ejection from a nucleus. Such intermediate states

can be detected in observable physical effects. For example, in the Coulomb field of the nucleus, *polarization of vacuum* as it were occurs, that is, a resultant displacement of the "background" due to pair production and annihilation. That is why the field acting on an electron close to the nucleus (of the order $h/(mc) \sim 10^{-11}$ cm) is not strictly a Coulomb field. This affects the configuration of the energy levels of atomic electrons (see further on).

Thus, the Dirac equation led to much more than a simple refinement of quantum mechanics in the relativistic domain. The concept of field was extended to particles, which led to the prediction of antiparticles.

Charge Conjugation. From what has been said of the positron one could gain the impression that it was something of an entirely different nature than the electron: the electron is a particle, the positron is a "hole". An apparent asymmetry developed between the charges of the two signs. However, a theoretical formulation is possible which completely restores the symmetry between the electron and the positron.

As said before, Eq. (37.11) allows for solutions corresponding to both positive and negative energies. Furthermore, to each energy sign there corresponds two spin projection signs, making for a total of four solutions. Those corresponding to the positive sign of energy have physical meaning. A filled background was added to eliminate the real appearance of the second two solutions, which have no natural confirmation.   .

The Dirac equation can be so transformed that the equation describing the positron becomes entirely the same as the equation for the electron. We are now speaking of a positron, that is, of an independent particle of positive energy, not a "hole". At the same time, such theoretical predictions as electron-positron production and annihilation and the polarization of vacuum remain valid, because the equations are from the outset written for fields, not separate particles.

Let us consider a transformation of the Dirac equation which leads to the wave function of the positron.

If the Lagrangian is varied with respect to the ψ^* components rather than the ψ components, the expression (37.8) yields Dirac's equation for a complex conjugate function. In this equation it is convenient to first write the wave function on the left of the operators, as it appears in the Lagrangian. In performing the variation we must also make a transformation by parts, mindful that L itself appears under the sign of a four-dimensional integral in the action expression, $S = \int L \, d^4\tau$. In this way the derivatives involved in p_t and **p** are transferred from the variations to the required functions,

with p_t accordingly being replaced by $-p_t$, and $\mathbf{p}$ by $-\mathbf{p}$. In the complex conjugate equation these operators operate not from left to right but, by definition, from right to left:

$$\psi^* \left(-p_t + \hat{\alpha}\mathbf{p} - mc\hat{\beta} \right) = 0 \tag{37.20a}$$

The operators $\hat{\alpha}$ and $\hat{\beta}$ also operate from right to left, that is, the kth component of the wave function, ψ_k^*, is multiplied by the matrix element in the kth row rather than the matrix element in the kth column. In equation form it looks like this:

$$\sum_{h=1}^{4} \alpha_{ih}\psi_h = (\hat{\alpha}\psi)_i, \qquad \sum_{h=1}^{4} \psi_h^* \alpha_{hi} = (\psi^*\hat{\alpha})_i \tag{37.21}$$

Consequently, to return to the usual notation, the columns and rows of the matrices $\hat{\alpha}$ and $\hat{\beta}$ must be interchanged. They are then denoted as $\widetilde{\hat{\alpha}}$ and $\widetilde{\hat{\beta}}$, and Eq. (37.20a) becomes

$$\left(-p_t + \widetilde{\hat{\alpha}}\hat{\mathbf{p}} - mc\widetilde{\hat{\beta}} \right)\psi^* = 0 \tag{37.20b}$$

But there also exists a transformation which reverts equation (37.21) completely to the initial form of the Dirac equation (37.11). Let us obtain this transformation, denoted by the letter C. It should be applied as an operator to Eq. (37.20a) on the left, requiring that as a result of the commutations with the operators $\hat{\alpha}$ and $\hat{\beta}$ we obtain an equation for the function $C\psi^*$ identical in form with (37.11). In other words, C must satisfy the following commutation relations:

$$C\widetilde{\hat{\alpha}} = \hat{\alpha}C \tag{37.22}$$

$$C\widetilde{\hat{\beta}} = -\hat{\beta}C \tag{37.23}$$

The specific form of C depends upon the choice of the matrices $\hat{\alpha}$ and $\hat{\beta}$. We assume that they satisfy the relation (37.10b). The matrices $\hat{\rho}$ and $\hat{\sigma}$ are Hermitian; hence, if they are made up of real elements, an interchange of rows and columns alters nothing. If they consist of purely imaginary elements, like the elements of $\hat{\rho}_2$ and $\hat{\sigma}_y$, an interchange reverses the sign of the matrix. Now, substituting in place of the components $\hat{\alpha}_x$, $\hat{\alpha}_y$, and $\hat{\alpha}_z$ and β their expressions (37.10b), we rewrite (37.22) and (37.23) in the form of equations in components:

$$C\hat{\rho}_1\hat{\sigma}_x = \hat{\rho}_1\hat{\sigma}_x C, \qquad C\hat{\rho}_1\hat{\sigma}_y = -\hat{\rho}_1\hat{\sigma}_y C$$

$$C\hat{\rho}_1\hat{\sigma}_z = \hat{\rho}_1\hat{\sigma}_z C, \qquad C\hat{\rho}_3 = -\hat{\rho}_3 C \tag{37.24}$$

It follows from this that

$$C = \hat{\rho}_2 \hat{\sigma}_y \qquad (37.25)$$

Since $\hat{\rho}_2$ is anticommutative with $\hat{\rho}_1$, and $\hat{\sigma}_y$ is anticommutative with $\hat{\sigma}_x$, the first equation in (37.24) holds. The other equations are verified in the same way.

Thus, operating with C on Eq. (37.21) and interchanging it with the operators $\widetilde{\alpha}$ and $\widetilde{\beta}$, we obtain

$$(p_t - \hat{\alpha}\hat{\mathbf{p}} - mc\hat{\beta})\, C\psi^* = 0 \qquad (37.26)$$

which is fully identical with (37.11). But the complex conjugate function ψ^* depends upon the coordinates and time according to the law

$$\psi^* = \psi^*(0)\, e^{iEt/h - i\mathbf{pr}/h} \qquad (37.27)$$

To it also correspond two energy signs: $E = \pm (c^2 p^2 + m^2 c^4)^{1/2}$. If we substitute energy with the negative sign into (37.27), reverse the direction of the momentum $\mathbf{p}$, and subject $\psi^*(0)$ to the C transformation, we obtain a wave function of a particle of positive energy satisfying the same equation as the electron. The momentum sign is taken the reverse that of the electron so as to have the same spatio-temporal dependence of the wave function.

We have thus proved that the function $C\psi^*$ can be considered as belonging to a positive electron (according to the sign of the momentum) with positive energy. In other words, $C\psi^*_E(-\mathbf{p})$ is a function of a positron moving in a field in the opposite direction of an electron.

The C transformation effects the transition from an electron to a positron. But since $C^2 = 1$, the same transformation transforms "positron" equations into "electron" equations, which establishes the symmetry between both particles.

The C transformation is known as *charge conjugation*, it "transfers" particles into antiparticles.

W. Pauli and V. F. Weisskopf showed that particles without spin can also have antiparticles. Subsequently such a particle was in fact discovered: the π-meson; π_+- and π_--mesons possess particle-antiparticle properties.

If the C transformation does not alter the form of the wave function of a particle, then the particle and antiparticle are identical. Such particles are at present termed absolutely neutral, as distinct, for example, from the electrically neutral neutron, which nevertheless has its antiparticle. Absolutely neutral are the π_0-meson and the light quantum.

The Transformation to the Nonrelativistic Wave Equation.
It is instructive to compare the relativistic wave equation with the
Schrödinger equation with the help of a limiting process. Since we
are primarily interested in the motion of an electron in an external
field, we first write the corresponding Dirac equation for an electron
interacting with an electromagnetic field. For this we replace the
momentum $\mathbf{p}$ by $\mathbf{p} - (e/c)\,\mathbf{A}$, and the energy E by $E + e\varphi$ (Sec. 14).
Thus, Dirac's equation for this case takes the form

$$p_t\psi = \hat{\boldsymbol{\alpha}}\left(\hat{\mathbf{p}} - \frac{e}{c}\mathbf{A}\right)\psi + mc\hat{\beta}\psi + e\varphi\psi \qquad (37.28)$$

We assume that the operators $\hat{\boldsymbol{\alpha}}$ and $\hat{\beta}$ have been chosen according
to (37.10b). We expand only $\hat{\boldsymbol{\rho}}$, and not $\hat{\boldsymbol{\sigma}}$. In other words, we treat
the first pair (ψ_1, ψ_2) and the second pair (ψ_3, ψ_4) as one entity (a
bispinor):

$$\chi_1 = \begin{pmatrix} \psi_1 \\ \psi_2 \end{pmatrix}, \qquad \chi_2 = \begin{pmatrix} \psi_3 \\ \psi_4 \end{pmatrix} \qquad (37.29)$$

We know how $\hat{\rho}_1$ and $\hat{\rho}_3$ operate on them. Writing this down in
explicit form, we obtain Eq. (37.28) for χ_1 and χ_2:

$$\frac{E}{c}\chi_1 = p_t\chi_1 = \hat{\boldsymbol{\sigma}}\left(\hat{\mathbf{p}} - \frac{e}{c}\mathbf{A}\right)\chi_2 + mc\chi_1 + \frac{e\varphi}{c}\chi_1$$

$$\frac{E}{c}\chi_2 = p_t\chi_2 = \hat{\boldsymbol{\sigma}}\left(\hat{\mathbf{p}} - \frac{e}{c}\mathbf{A}\right)\chi_1 - mc\chi_2 + \frac{e\varphi}{c}\chi_2 \qquad (37.30)$$

From the second equation it follows that

$$\chi_2 = \frac{1}{E/c + mc - e\varphi/c}\,\hat{\boldsymbol{\sigma}}\left(\hat{\mathbf{p}} - \frac{e}{c}\mathbf{A}\right)\chi_1 \qquad (37.31)$$

In nonrelativistic motion, when $v \ll c$, the energy E of the particle
is very close to mc^2, so that the whole denominator in (37.31) is
in the first, initial approximation replaced by $2mc$. Now, substitut-
ing χ_2 into the first equation of (37.30), we find that the two-compo-
nent function χ_1 satisfies the following equation:

$$(E - mc)\chi_1 = (p_t - mc)\chi_1$$
$$= \frac{1}{2mc}\left[\hat{\boldsymbol{\sigma}}\cdot\left(\hat{\mathbf{p}} - \frac{e}{c}\mathbf{A}\right)\right]^2\chi_1 + \frac{e\varphi}{c}\chi_1 \qquad (37.32)$$

The difference $E - mc$ in the left-hand side of the equation is
the nonrelativistic Hamiltonian of the particle, $\hat{\mathcal{H}}$, divided by c.
We transform the first expression attached to χ_1 in the right-hand side
of (37.32), which, after multiplying by $2mc$, we write as follows:

$$\left[\hat{\boldsymbol{\sigma}}\cdot\left(\hat{\mathbf{p}} - \frac{e}{c}\mathbf{A}\right)\right]^2 = (\boldsymbol{\sigma}\cdot\hat{\mathbf{p}})^2 + \frac{e^2}{c^2}(\hat{\boldsymbol{\sigma}}\cdot\mathbf{A})^2$$

$$- \frac{e}{c}[(\hat{\boldsymbol{\sigma}}\hat{\mathbf{p}})(\hat{\boldsymbol{\sigma}}\cdot\mathbf{A}) + (\hat{\boldsymbol{\sigma}}\cdot\mathbf{A})(\hat{\boldsymbol{\sigma}}\hat{\mathbf{p}})] \qquad (37.33)$$

Here the first term is

$$(\hat{\boldsymbol{\sigma}}\hat{\mathbf{p}})^2 = \hat{\sigma}_x^2\hat{p}_x^2 + \hat{\sigma}_y^2\hat{p}_y^2 + \hat{\sigma}_z^2\hat{p}_z^2 + (\hat{\sigma}_x\hat{\sigma}_y + \hat{\sigma}_y\hat{\sigma}_x)\,\hat{p}_x\hat{p}_y$$
$$+ (\hat{\sigma}_x\hat{\sigma}_z + \hat{\sigma}_z\hat{\sigma}_x)\,\hat{p}_x\hat{p}_z + (\hat{\sigma}_y\hat{\sigma}_z + \hat{\sigma}_z\hat{\sigma}_y)\,\hat{p}_y\hat{p}_z$$
$$= \hat{p}_x^2 + \hat{p}_y^2 + \hat{p}_z^2 = \hat{p}^2 \tag{37.34}$$

and similarly

$$(\hat{\boldsymbol{\sigma}}\cdot\mathbf{A})^2 = \mathbf{A}^2 \tag{37.35}$$

Making use of the anticommutativity of the components of $\hat{\boldsymbol{\sigma}}$, we reduce the last term in (37.33) to the form

$$(\hat{\boldsymbol{\sigma}}\cdot\hat{\mathbf{p}})\,(\hat{\boldsymbol{\sigma}}\cdot\mathbf{A}) + (\hat{\boldsymbol{\sigma}}\cdot\mathbf{A})\,(\hat{\boldsymbol{\sigma}}\cdot\hat{\mathbf{p}})$$
$$= (\mathbf{A}\cdot\hat{\mathbf{p}}) + (\hat{\mathbf{p}}\cdot\mathbf{A}) + [(\hat{p}_xA_y - A_y\hat{p}_x) - (\hat{p}_yA_x - A_x\hat{p}_y)]\,i\hat{\sigma}_z\,^\bullet$$
$$+ [(\hat{p}_zA_x - A_x\hat{p}_z) - (\hat{p}_xA_z - A_z\hat{p}_x)]\,i\hat{\sigma}_y$$
$$+ [(\hat{p}_yA_z - A_z\hat{p}_y) - (\hat{p}_zA_y - A_y\hat{p}_z)]\,i\hat{\sigma}_x \tag{37.36}$$

where the properties of the Pauli matrices, (30.34a)-(30.34c), have been used.

The commutator $\hat{p}_xA_y - A_y\hat{p}_x$ is equal to

$$\frac{h}{i}\left(\frac{\partial}{\partial x}A_y - A_y\frac{\partial}{\partial x}\right) = \frac{h}{i}\frac{\partial A_y}{\partial x} \tag{37.37}$$

so that the factor of $\hat{\sigma}_z$ is the z component of the magnetic field, H_z:

$$\left(\frac{\partial A_y}{\partial x} - \frac{\partial A_x}{\partial y}\right) = H_z$$

The same holds for the other two components of $\mathbf{H}$.

Collecting all the terms in (37.32), we obtain

$$\hat{\mathcal{H}}\chi_1 = \frac{1}{2m}\left(\hat{\mathbf{p}} - \frac{e}{c}\mathbf{A}\right)^2\chi_1 + e\varphi\chi_1 - \frac{eh}{2mc}(\hat{\boldsymbol{\sigma}}\cdot\mathbf{H}) \tag{37.38}$$

The first two terms in the right-hand side of (37.38) represent the nonrelativistic Hamiltonian of a spinless particle in an external electromagnetic field. The last term is the energy of the magnetic moment

$$\hat{\boldsymbol{\mu}} = \frac{eh}{2mc}\hat{\boldsymbol{\sigma}} \tag{37.39}$$

in the external magnetic field, $\mathbf{H}$. Thus, from the Dirac equation we have obtained the correct relationship between the magnetic moment and the angular momentum of an electron, (30.51).

Also of interest is the development of a formula for the interaction energy between the magnetic spin and orbital angular momenta of

the electron. This quantity involves the square of the speed of light in the denominator, and is therefore obtained in the higher than (37.38) approximation. We, however, shall not seek all the terms of this approximation and take only what is of interest to us. We can see from Eq. (37.31) that the higher-order term in c^{-2} is

$$\chi_2' = \frac{e\varphi}{4m^2c^2}\, \hat{\sigma} \left(\hat{\mathbf{p}} - \frac{e}{c}\, \mathbf{A} \right) \chi_1 \tag{37.40}$$

Substituting the correction (37.40) into (37.30), we see that an operator involving both the electric field and the spin can only be obtained from the expression

$$(\hat{\sigma} \cdot \hat{\mathbf{p}})\, e\varphi\, (\hat{\sigma} \cdot \hat{\mathbf{p}}) \tag{37.41}$$

Let us interchange the potential, φ, with the second factor $(\hat{\sigma} \cdot \hat{\mathbf{p}})$. Using a formula similar to (37.37), we can write

$$e\varphi\, (\hat{\sigma} \cdot \hat{\mathbf{p}}) = (\hat{\sigma} \cdot \hat{\mathbf{p}})\, e\varphi - \frac{h}{i}\, (\hat{\sigma} \cdot \mathrm{grad}\ \varphi) = (\hat{\sigma} \cdot \hat{\mathbf{p}})\ e\varphi - ih\, (\hat{\sigma} \cdot \mathbf{E})$$

The product $(\hat{\sigma} \cdot \hat{\mathbf{p}})\, (\hat{\sigma} \cdot \mathbf{E})$ expands according to the general rule (30.36):

$$(\hat{\sigma} \cdot \hat{\mathbf{p}})\, (\hat{\sigma} \cdot \mathbf{E}) = (\hat{\mathbf{p}} \cdot \mathbf{E}) + i\hat{\sigma}\, (\hat{\mathbf{p}} \times \mathbf{E})$$

Here, the first term does not depend upon the spin and is immaterial for the spin-orbit interaction. In the second term we make use of the fact that the self-consistent field acting on the electron is a central field, so that $\mathbf{E} = -(\mathbf{r}/r)(d\varphi/dr)$, and $\mathbf{r} \times \hat{\mathbf{p}}$ is the orbital angular momentum operator of the electron, $\hat{\mathbf{l}}$. Expressing it in terms of units of h, we obtain the formula for the energy operator of the spin-orbit interaction:

$$\hat{V} = \frac{eh^2}{4m^2c^2}\, (\hat{\sigma} \cdot \hat{\mathbf{l}})\, \frac{1}{r}\, \frac{d\varphi}{dr} \tag{37.42}$$

Equation (37.38) is applicable only to the electron. Although the proton also has half-integral spin, its magnetic moment is 2.9 times greater than obtained in the last term of (37.38). The neutron also possesses magnetic moment, equal, in the same units, to 1.9. At the same time, according to Dirac's theory a neutral particle should not have any magnetic moment at all, since e in (37.38) denotes charge.

Usually the following explanation is offered for the reason why the proton and the neutron are not subject to the Dirac equation. Both nuclear particles interact very strongly via a meson field. For that reason they continuously emit and absorb mesons, in the same way as an electromagnetic field produces electron-positron pairs which immediately annihilate. But whereas such pairs do not make a great contribution to the total electromagnetic effect (since

the measure of charge-field interaction is a small quantity $e^2/(hc) = 1/137$), nuclear interactions (which are very strong) make it impossible to consider neutrons and protons as isolated from the meson field surrounding them.

This explanation has not been verified by quantitative computations, because the theory of nuclear forces has not yet been developed. However, a direct study of the electromagnetic field of the proton and the neutron with very fast electrons indicates that both are indeed surrounded by charges and currents in a region of about 10^{-14} cm.

The Fine Structure Formula. As an exception, we shall present without proof an important formula for the energy eigenvalues of an electron in the hydrogen atom or in any Coulomb field of charge Ze:

$$\frac{E}{mc^2} - 1 = \left(1 + \frac{\alpha^2 Z^2}{\{n - (j + 1/2) + [(j + 1/2)^2 - \alpha^2 Z^2]^{1/2}\}^2} \right)^{-1/2}$$

(37.43)

Here, n is the principal quantum number, $j = l \pm 1/2$ is the total angular momentum of the electron, and $\alpha = e^2 (hc)$. If Ze is assumed small in comparison with unity, we obtain the nonrelativistic formula (29.41).

Equation (37.43) agrees with the result of Exercise 9, Section 14, in a remarkable way. If in the classical formula we replace the action variables in terms of Bohr's quantum conditions (31.42), we obtain Eq. (37.43), which was developed in this way by Sommerfeld without taking electron spin into account. But in that case, of course, the number of atomic states comes out wrong.

Radiation Corrections. It follows from Eq. (37.43) that the energies of states $2s_{1/2}$ and $2p_{1/2}$ must be the same, since they correspond to the same n and j. Actually, though, they vary in terms of frequency by 1043 megacycles per second.

The energy difference is due to the fact that in developing the fine structure formula (37.43) the effect of zero-point oscillations of the electromagnetic-field vacuum was not taken into account. These oscillations differently affect the electron in the s and p states and therefore split the degenerate energy level.

Besides zero-point oscillations, a certain contribution to the splitting is made by the already mentioned polarization of vacuum due to the production and annihilation of electron-positron pairs. Since the polarization of vacuum occurs mainly at small distances from the nucleus (of the order of $h/(mc)$), while the electron in a hydrogen atom is much farther away, at a distance of the order of one atomic unit, that is $h^2/(me^2)$, the polarization contribution to the splitting of the energy level is relatively small: around three

per cent of the total effect. Nevertheless, the precision of experimental data is so great that the reality of the effect of vacuum polarization has been fully confirmed.

The corrections to the formulas due to zero-point oscillations or pair production and annihilation are known as *radiation corrections*.

In calculating them diverging integrals always appear. Therefore the following procedure is adopted. The energy shift of a free electron due, say, to zero-point oscillations is calculated, together with the same shift of an electron bound in an atom. Both corrections lead to diverging integrals, but their difference can be determined, because it is finite; moreover, this process is quite unambiguous.

The meaning of this subtraction consists in the following. Physically the electron is inseparable from its charge, that is from the radiation field. When we speak of a "free" electron, we actually have in mind that the electron interacts with the radiation field which, as was shown, cannot be assumed zero even in the ground state. Thus, by subtracting the energy corrections for a free electron from the energy corrections for a bound electron, we thereby simply redefine the concept of a free electron.

Ultimately a small and finite shift is obtained, the relative smallness of which is due to the fact that the *fine structure constant*, $e^2/(hc)$, is small in comparison with unity.

Similarly, we are able to find the correction to the magnetomechanical ratio of the electron, that is $eh/(2mc)$. It agrees with experiments in the next two approximations with respect to $e^2/(hc)$.

Thus, quantum electrodynamics meets the basic requirements of any physical theory: it can precalculate any observable effect, to any degree of accuracy, uniquely, and intrinsically unambiguously, in full agreement with experiments.

The most important unsolved problem of quantum electrodynamics consists in the theoretical development of the dimensionless quantity 1/137. So far, however, we do not even know if this is at all possible in the framework of electrodynamics alone, without introducing other fields besides the electromagnetic.

EXERCISES

1. Prove that a quantum cannot give rise to an electron-positron pair in free space in the absence of an additional external field.

Solution. The conservation laws in the absence of a field are written thus:

$$-(m^2c^4 + c^2p^2)^{1/2} + h\omega = (m^2c^4 + c^2p_1^2)^{1/2}, \qquad \mathbf{p} + \frac{h\omega}{c}\,\mathbf{n} = \mathbf{p_1}$$

Here, **p** is the electron momentum in a negative energy state, **n** is the unit vector in the direction of the quantum momentum, $\mathbf{p}_1$ is the electron momentum in a positive energy state. Substituting $\mathbf{p}_1$ in the first equality and squaring the left- and right-hand sides, it is easy to see that this equation is not satisfied.

Another method of proof is based on simple reasoning. A transition to another inertial reference frame can always make the energy of a quantum less than $2mc^2$. A quantum cannot give rise to a pair in such a reference frame, simply because it has insufficient energy. But what is impossible in one reference frame is impossible in all reference frames, because the possibility or impossibility of an event does not depend upon the choice of reference frame.

The preceding argument no longer holds if pair production is considered close to a nucleus. Here, the nucleus is at rest in one reference frame and in motion in another. Where the energy of the quantum is less than $2mc^2$, the moving nucleus will "help" it produce a pair. Naturally, pair production is impossible if the energy of the electron in the rest frame of the nucleus is less than $2mc^2$.

2. Obtain the solution to the Dirac equation for a free electron.

Solution. Equate ψ_1 to zero. Then the first equation of (37.11) is satisfied if we take $\psi_3 = Ac\,(p_x - ip_y)$ and $\psi_4 = -Acp_z$. Here $\hat{\alpha}$ and $\hat{\beta}$ are determined by (37.10b), and not by (37.10a). The second equation of (37.11) gives

$$\psi_2 = \frac{Ac^2\,(p_x^2 + p_y^2 + p_z^2)}{E - mc^2} = \frac{A\,(E^2 - m^2c^4)}{E - mc^2} = A\,(E + mc^2)$$

The third equation of (37.11) reduces to the identity

$$(E + mc^2)\,\psi_3 = Ac\,(E + mc^2)\,(p_x - ip_y)$$
$$= c\,(p_x - ip_y)\,\psi_2 = Ac\,(p_x - ip_y)\,(E + mc^2)$$

The fourth equation also reduces to an identity. The number A is determined from the normalization condition $|\psi_1|^2 + |\psi_2|^2 + |\psi_3|^2 + |\psi_4|^2 = 1$, or $A = [2E\,(E + mc^2)]^{-1/2}$.

The components ψ_3 and ψ_4 are small compared with ψ_1, ψ_2 if $v \ll c$. Therefore the solution corresponds to positive energy. Another solution with positive energy is obtained if we take $\psi_2 = 0$. Negative energy solutions are obtained if we choose $\psi_3 = 0$ or $\psi_4 = 0$.

3. Show that from the Dirac equation there follows a charge–conservation equation which is analogous to (23.15):

$$\frac{\partial}{\partial t}\,|\psi|^2 = -\operatorname{div}\,(\psi^* c\hat{\alpha}\psi)$$

where $|\psi|^2 = |\psi_1|^2 + |\psi_2|^2 + |\psi_3|^2 + |\psi_4|^2$.

Hint. Write Eq. (37.11) and its complex conjugate; multiply the first by ψ^* and the second by ψ; subtract the second from the first and utilize the Hermiticity of $\hat{\alpha}$ and $\hat{\beta}$.

4. Show that if ψ is a solution with positive energy E, then $\hat{\rho}_2\psi$ is a solution with negative energy $-E$.

Solution. The equation for ψ is

$$E\psi = c\,(\hat{\boldsymbol{\alpha}}\cdot\hat{\mathbf{p}})\,\psi + mc^2\hat{\beta}\psi$$

Whence

$$E\hat{\rho}_2\psi = c\hat{\rho}_2\,(\hat{\boldsymbol{\alpha}}\cdot\hat{\mathbf{p}})\,\psi + mc^2\hat{\rho}_2\hat{\beta}\psi = -[c\,(\hat{\boldsymbol{\alpha}}\cdot\hat{\mathbf{p}}) + mc^2\hat{\beta}]\,\hat{\rho}_2\psi$$

This proves that a negative-energy solution cannot be avoided.

5. Prove that the operators

$$\frac{h}{2}\,\hat{\sigma}_x = \frac{h}{2i}\,\hat{\alpha}_y\hat{\alpha}_z, \quad \frac{h}{2}\,\hat{\sigma}_y = \frac{h}{2i}\,\hat{\alpha}_z\hat{\alpha}_x, \quad \frac{h}{2}\,\hat{\sigma}_z = \frac{h}{2i}\,\hat{\alpha}_x\hat{\alpha}_y$$

operating on four-component functions are spin moment operators.

Solution. Since

$$\frac{h^2}{4}\,\hat{\sigma}_x^2 = -\frac{h^2}{4}\,\hat{\alpha}_y\hat{\alpha}_z\hat{\alpha}_y\hat{\alpha}_z = \frac{h^2}{4}\,\hat{\alpha}_y^2 = \frac{h^2}{4}$$

$$\frac{h^2}{4}\,\hat{\sigma}_x\hat{\sigma}_y = -\frac{h^2}{4}\,\hat{\alpha}_y\hat{\alpha}_z\hat{\alpha}_z\hat{\alpha}_x = -\frac{h^2}{4}\,\hat{\alpha}_y\hat{\alpha}_x = i\,\frac{h}{2}\,\hat{\sigma}_z$$

$$\frac{h^2}{4}\,\hat{\sigma}_y\hat{\sigma}_x = -i\,\frac{h}{2}\,\hat{\sigma}_z$$

the spin operators determined here possess all the required properties (see Sec. 30). This can also be seen from the definition (37.10*b*) of $\hat{\boldsymbol{\alpha}}$ in terms of $\hat{\boldsymbol{\sigma}}$ and $\hat{\boldsymbol{\rho}}$. We notice that the spin operators do not interchange both functions of the first pair (ψ_1, ψ_2) and both functions of the second pair (ψ_3, ψ_4) but make interchanges only inside each pair.

6. Show that according to the Dirac equation only the sum of the orbital and spin angular momenta and not each angular momentum separately satisfies the angular-momentum conservation law.

Solution. The total angular momentum is defined as

$$\hat{\mathbf{j}} = \hat{\mathbf{l}} + \hat{\mathbf{s}} = \hat{\mathbf{r}}\times\hat{\mathbf{p}} + \frac{h}{2}\,\hat{\boldsymbol{\sigma}}$$

$$\hat{j}_z = \hat{l}_z + \hat{s}_z = x\hat{p}_y - y\hat{p}_x + \frac{h}{2i}\,\hat{\alpha}_x\hat{\alpha}_y$$

We calculate the commutator with the Hamiltonian:

$$\hat{\mathcal{H}}\hat{j}_z - \hat{j}_z\hat{\mathcal{H}}$$

$$= [c\,(\hat{\alpha}_x\hat{p}_x + \hat{\alpha}_y\hat{p}_y + \hat{\alpha}_z\hat{p}_z) + \hat{\beta}mc^2]\left(x\hat{p}_y - y\hat{p}_x + \frac{h}{2i}\,\hat{\alpha}_x\hat{\alpha}_y\right)$$

$$- \left(x\hat{p}_y - y\hat{p}_x + \frac{h}{2i}\,\hat{\alpha}_x\hat{\alpha}_y\right)[c\,(\hat{\alpha}_x\hat{p}_x + \hat{\alpha}_y\hat{p}_y + \hat{\alpha}_z\hat{p}_z) + \hat{\beta}mc^2]$$

$$= c\hat{\alpha}_x\hat{p}_y\,(\hat{p}_x x - x\hat{p}_x) - c\hat{\alpha}_y\hat{p}_x\,(\hat{p}_y y - y\hat{p}_y)$$

$$+ \frac{hc}{2i}\,\hat{p}_x\,(\hat{\alpha}_x\hat{\alpha}_x\hat{\alpha}_y - \hat{\alpha}_x\hat{\alpha}_y\hat{\alpha}_x) + \frac{hc}{2i}\,\hat{p}_y\,(\hat{\alpha}_y\hat{\alpha}_x\hat{\alpha}_y - \hat{\alpha}_x\hat{\alpha}_y\hat{\alpha}_y)$$

$$= \frac{hc}{i}\,(\hat{\alpha}_x\hat{p}_y - \hat{\alpha}_y\hat{p}_x + \hat{p}_x\hat{\alpha}_y - \hat{p}_y\hat{\alpha}_x) = 0$$

The Hamiltonian commutes also with the square of the total angular momentum $\hat{j}^2 = \hat{j}_x^2 + \hat{j}_y^2 + \hat{j}_z^2$, which is proved similarly. The integrals of motion are $\hat{j}^2$ and $\hat{j}_z$, and not $\hat{l}^2$, $\hat{l}_z$ and $\hat{s}^2$, $\hat{s}_z$ separately.

7. Show that the Dirac equation is invariant with respect to the substitution of $-t$ for t, that is, to the time reversal operation (the T operator).

Solution. Replacement of t by $-t$ corresponds to a transformation to the complex conjugate equation $\psi^*(-p_t - \widetilde{\widehat{\alpha}}\,\widetilde{\mathbf{p}} - \widetilde{\beta}mc^2) = 0$. We must find the transformation that reverses it to the initial form of the Dirac's equation. We pass to the transposed operators and get

$$(-p_t - \widetilde{\widehat{\alpha}}\,\hat{\mathbf{p}} - \widetilde{\beta}mc^2)\,\psi^* = 0$$

(see (37.20b)). The required transformation must satisfy the conditions

$$T\widetilde{\widehat{\alpha}} = -\hat{\alpha}T, \qquad T\widetilde{\beta} = -\hat{\beta}T$$

or

$$T\hat{\alpha}_x = -\hat{\alpha}_x T, \quad T\hat{\alpha}_y = -\hat{\alpha}_y T, \quad T\hat{\alpha}_z = -\hat{\alpha}_z T, \quad T\hat{\beta} = -\hat{\beta}T$$

From this, $T = \hat{\rho}_1$, $\hat{\sigma}_y = \hat{\alpha}_y$. Consequently $CPT = i'$, that is, it does not alter the Dirac equation.

SUPPLEMENTARY EXERCISES

Sec. 1. (a) What is the unit of mass in grams, taking as the standard the gravitational constant G in the gravitation law $F = Gm_1m_2/r^2$, the centimetre, and the second?

(b) What path relative to the earth is described by a person walking uniformly along the radius of a uniformly rotating platform?

Sec. 2. (a) How many degrees of freedom do the following objects have: a pair of scissors? A bicycle (not counting the chain)? A meat grinder attached to a table?

(b) Offer an example of motion of a system in conditions having the same symmetry as a sphere or a cylinder.

Sec. 3. Write the Lagrange equations for a system with the Lagrangian

$$L = \frac{m}{2}\,(\dot{q}_1^2 + q_1^2\dot{q}_2^2 + q_1^2\sin^2 q_2\,\dot{q}_3^2) - U\,(q_1)$$

Sec. 4. (a) The proton-neutron binding energy in a deuterium nucleus is 2.2 MeV. To what energy must a proton be accelerated to smash a deuterium nucleus? Assume the masses of the proton and the neutron to be equal

(b) What is the energy distribution between shell and gun at the time of firing?

Sec. 5. (a) What is the ratio of the angular momenta of two planets of equal mass orbiting the sun along circular orbits of radii r_1 and r_2?

(b) Express the major axis of an elliptical orbit in terms of the energy of the planet.

(c) Express the period of rotation of a planet in terms of its energy, making use of the fact that for $\varepsilon < 1$

$$\int_0^{2\pi} d\varphi\,(1 + \varepsilon \cos \varphi)^{-2} = 2\pi\,(1 - \varepsilon^2)^{-3/2}$$

Sec. 6. (a) Determine the mean energy transferred from a colliding particle to an identical particle initially at rest, for the case of isotropic scattering in the centre-of-mass reference frame.

(b) Determine the maximum angle at which a particle of mass $3m$ may be scattered on a resting particle of mass m.

(c) A particle of mass m_1 impinges on a particle of mass m_2, which is at rest, and adheres to it. Express the kinetic energy of the combined particle in terms of the kinetic energy of the impinging particle.

Sec. 7. (a) Consider the oscillations of a particle with one degree of freedom in the presence of an elastic force $-m\omega_0^2 x$ and the force friction directed against the velocity and proportional to its magnitude.

Solution. From the condition, the equation of motion is

$$m\ddot{x} = -m\omega_0^2 x - \alpha\dot{x}$$

The solution has the form $x = \mathrm{Re}\,(Ce^{-i\omega t})$, where C and ω are complex quantities. For ω we obtain the quadratic equation $m\omega^2 - m\omega_0^2 + i\alpha\omega = 0$, whence

$$\omega = -\frac{i\alpha}{2m} \pm \left(\omega_0^2 - \frac{\alpha^2}{4m^2}\right)^{1/2}$$

If $\omega_0 > \alpha/(2m)$, the square root is real, and the solution has the form of damped oscillations:

$$x = e^{-\alpha t/(2m)}\,\mathrm{Re}\left\{C\exp\left[\pm i\left(\omega_0^2 - \frac{\alpha^2}{4m^2}\right)^{1/2}\right]\right\}$$

If the friction is great, the root is imaginary and

$$\omega = -i\left[\frac{\alpha}{2m} \mp \left(\frac{\alpha^2}{4m^2} - \omega_0^2\right)^{1/2}\right]$$

The solution is aperiodic and falls off exponentially:

$$x = e^{-\alpha t/(2m)}\left\{C_1\exp\left[-\left(\frac{\alpha^2}{4m^2} - \omega_0^2\right)^{1/2}t\right] + C_2\exp\left[\left(\frac{\alpha^2}{4m^2} - \omega_0^2\right)^{1/2}t\right]\right\}$$

(b) Consider the motion of a particle in the presence of an elastic force and an external driving force $\mathrm{Re}\,(fe^{-i\omega t})$, where $f \equiv |f|\,e^{-i\zeta}$.

Answer.

$$x = \text{Re}\left(\frac{fe^{-i\omega t}}{m(\omega_0^2 - \omega^2)}\right) + \text{Re}\left(Ce^{-i\omega_0 t}\right)$$

(c) Ditto, for the case $\omega = \omega_0$ (resonance).

Answer. The term due to the external force has the form

$$x = \text{Re}\left(\frac{fte^{-i\omega_0 t}}{-2im\,\omega_0}\right)$$

The oscillation amplitude increases linearly with time.

(d) Solve problem (b) on the assumption that in addition to the driving force there is a frictional force $-\alpha \dot{x}$.

Answer. The term due to the external force is

$$x = \text{Re}\left(\frac{fe^{-i\omega t}}{m(\omega_0^2 - \omega^2) - i\alpha\omega}\right) = \frac{|f|\cos(\omega t + \zeta - \psi)}{[m^2(\omega_0^2 - \omega^2)^2 + \alpha^2\omega^2]^{1/2}}$$

$$\psi = \text{arc tan}\,\frac{\alpha\omega}{m(\omega_0^2 - \omega^2)}$$

The oscillation amplitude remains finite at all frequency values of the driving force.

The phase difference ψ between the oscillations of the driving force and the mass point is for three different cases thus:

(i) $\omega \ll \omega_0$, $\alpha\omega/m \ll \omega_0^2$, $\psi = 0$. The particle follows the driving force, coming in step with its phase.

(ii) $\omega \gg \omega_0$, $\alpha/m \ll \omega$, $\psi = -\pi$. Owing to inertia the particle oscillates in the opposite phase with respect to the driving force.

(iii) $\omega = \omega_0$ (resonance), $\psi = -\pi/2$. The phase of the particle is shifted by $-\pi/2$ with respect to the phase of the driving force.

(e) Consider the oscillations of an elastically bound particle in the presence of "dry" friction F directed always against the velocity, and in magnitude independent of the velocity.

Solution. For the first half-period, assuming the velocity during it to be positive, we have the equation of motion $m\ddot{x} + m\omega_0^2 x = -F$. Multiply this equation by $\dot{x}$ and integrate from certain initial values of the coordinate and velocity, x_0 and $\dot{x}_0$. The integral of the motion reduces to the form

$$\frac{\dot{x}^2}{\omega_0^2} + \left(x + \frac{E}{m\omega_0^2}\right)^2 = \frac{\dot{x}_0^2}{\omega_0^2} + \left(x_0 + \frac{F}{m\omega_0^2}\right)^2$$

This is the equation of a semicircle in the plane $(x, \dot{x}/\omega_0)$ centred on $x = -F/(m\omega_0^2)$, $\dot{x} = 0$. The semicircle terminates at the intersection with the x axis, where the velocity reverses its sign. If at that point $x \leqslant F/(m\omega_0^2)$, the motion ceases, since the elastic force is no longer able to overcome the force of friction, which from that instant is in the opposite direction. If

$x > F/(m\omega_0^2)$ at $\dot{x} = 0$, the motion continues until the rest point occurs within the interval $-F/(m\omega_0^2) \leqslant x \leqslant F/(m\omega_0^2)$.

This phenomenon is called *stagnation*. It lowers the accuracy of instruments in which dry friction has not been eliminated.

Sec. 8. (a) At what latitude will the oscillation plane of a Foucault pendulum describe a complete circle in 48 hours?

(b) Would the oscillation plane of a pendulum rotate on the moon?

Sec. 9. (a) Determine the principal axes of a molecule of HDO, assuming the bond lengths of HO and DO to be equal and at an angle of 108° (D is deuterium).

(b) Determine the frequency of small oscillations of a homogeneous elliptical cylinder with semiaxes a and b lying in the horizontal plane.

Sec. 10. (a) Find the Hamiltonian if the Lagrangian is given in the problem to Section 3, and write the corresponding Hamilton equations.

(b) Write the Hamilton-Jacobi equation for the foregoing exercise and separate the variables.

(c) Write the expressions for the adiabatic invariants of a free symmetric top.

(d) Find the path of motion in Kepler's problem, assuming that the momentum projections are plotted along the coordinate axes.

Hint. As in the case of (10.13), determine the transformation function $V'(p, P)$, and write the Hamilton-Jacobi equation for it. Introduce the square of the momentum and its direction in the plane of motion as the dependent variables.

Sec. 11. (a) Write the components of curl of a certain vector in cylindrical coordinates.

(b) Compare the expressions for the Laplacians of a scalar and a vector in cylindrical coordinates.

Sec. 12. (a) Write Maxwell's equations and the equations for the potentials in spherical coordinates.

Sec. 13. (a) A particle travelling with a velocity $v_1 = 0.95c$ emits another particle, which travels relative to it with a velocity $v_2 = 0.99c$ at an angle of 30° in a reference frame in which the emitting particle rests. Find the magnitude and direction of the emitted particle with respect to the initial (fixed) frame of reference.

(b) A particle travelling with a velocity $v = 0.999c$ disintegrate into two particles of zero mass. Determine the angle between the directions of emission of the photons in the laboratory frame of reference, if in the reference frame fixed with respect to the disintegrating particle one of the zero-mass particles is travelling at a 30° angle to the direction of the velocity of the disintegrating particle.

Sec. 14. (a) Through what potential difference must an initially stationary electron pass so as to attain a velocity of $0.999c$?

(b) Determine the motion of a charge in constant and homogeneous electric and magnetic fields whose directions coincide.

(c) Ditto, in perpendicular fields.

(d) Show that the path of a charge in a constant and homogeneous magnetic field is of similar form in coordinate and momentum space (that is, a space in which the momentum components are plotted along the axes), and is not determined by the form of the energy-momentum dependence.

Hint. Compare with exercise (d) to Section 10.

Sec. 15. (a) A constant magnetic field acts parallel to the plates of a charged plane capacitor. Explain why no energy flux develops, although formally the Poynting vector is not zero.

Hint. Find the divergence of the Poynting vector.

(b) Using the relativistically invariant notation of the equations of motion of a charge in an electromagnetic field (14.33), show that in a constant, homogeneous field the solution has the form $x_i = \sum\limits_{n=1}^{4} A_i^{(n)} e^{\varkappa^{(n)}s}$, where the quantities $\varkappa^{(n)}$ are expressed in terms of the invariants of the electromagnetic field tensor.

(c) Show that the invariants of the tensor $F_{ik}^{*} = \varepsilon_{iklm}F_{lm}$ coincide with the invariants of the tensor F_{ik}.

Sec. 16. (a) A quadrupole is formed by four charges, 1, -1, 2, -2, located at the apexes of a parallelogram with sides 1 and 2 forming an angle 45°. Find the principal axes of the tensor of the quadrupole moment and the value of the moment relative to the principal axes.

(b) The principal moments of inertia of an ellipsoid of rotation are I_1, $I_2 = I_3$. Determine, in the quadrupole approximation, the components of the potential of the force of gravity on the symmetry axis and in the median plane perpendicular to it.

(c) A diatomic molecule with moments of inertia $I_1 = 0$, $I_2 = I_3$, and dipole moment d is placed in a constant, homogeneous electric field. Write its equations of motion and reduce to quadratures (the integral cannot be expressed in elementary functions).

Sec. 17. A symmetric molecule with principal moments of inertia I_1, $I_2 = I_3$ and with the magnetic moment rigidly connected with the direction of the first principal axis of inertia, is placed in a constant magnetic field. Show that its motion is similar to the motion of a symmetric top in a gravitational field.

Sec. 18. (a) The electric field components of a plane wave travelling along the x axis are: $E_y = E_1 \cos \beta \pm E_2 \sin \beta$, $E_z = E_1 \sin \beta - E_2 \cos \beta$. Using (18.36), find the components of the complex vector **F** and show that (i) the relation (18.38) is identically satisfied, (ii) $F_1^2 + F_2^2 = E_1^2 + E_2^2$, and (iii) the absolute magnitude of vectors F_1 and F_2 and the angle between them do not depend upon angle β.

(b) At the initial time, the E_y component of the electric field of a plane electromagnetic wave travelling along the x axis was given by the function $f(x)$, and its time derivative by the function $g(x)$. Write the expression for E_y at an arbitrary instant.

(c) Compare the relations (18.21), (18.22), and (18.31) with the formula expressing the energy of a zero-mass particle in terms of its momentum.

Sec. 19. (a) A plane monochromatic wave of frequency ω impinges normal on a screen with an aperture of radius a. Determine the approximate size of the illuminated region on another screen parallel to the first at a distance $d \gg a$ beyond the aperture, and find the conditions at which the diffraction region is considerably larger than that obtained on the basis of geometrical optics, that is, from a construction of rectilinear rays.

Sec. 20. (a) A plane circularly-polarized wave impinges on an electron. Find the elliptical polarization of the wave scattered at angle θ to the initial direction.

(b) Taking into account that the resultant momentum of the electromagnetic field radiated by a dipole is zero, show that a charge at rest in the field of a plane electromagnetic wave is subject to a force $(2/3) e^2 |E|^2/(m^2 c^4)$ in the direction of propagation of the wave.

(c) A charge at rest is subject to the action of the electromagnetic field of a travelling electromagnetic wave varying according to the law $E = E_0 \sin \omega t$. Show that oscillations of double frequency take place in the perpendicular direction, and determine the energy radiated by the oscillations per unit time.

Sec. 21. (a) Assuming the phase velocity u to be a function of one Cartesian coordinate x, and using (21.7) as the Hamilton-Jacobi equation, find the equation of a light beam of given frequency ω.

(b) Find the relation between the phase velocity, group velocity, and velocity of light in vacuum for the case when $\omega = c(k_0^2 + k^2)^{1/2}$.

Sec. 22. Consider the following thought experiment offered to "refute" the uncertainty principle. We have a screen with two apertures, one above the other. Particles are passed through the apertures one by one. The diffraction pattern produced by them on another screen is such as though each particle had passed through both apertures in the first screen. At the same time, the vertical component of the momentum received by the second screen on impact of a particle is measured. The assertion is made that if this component is directed upwards, the particle must have passed through the lower aperture in the first screen, and vice versa. This assertion is incompatible with the uncertainty principle. Show the error in the reasoning.

Hint. Apply the uncertainty principle to the second screen as a universal principle and show that in measuring the momentum the uncertainty in the coordinate is equal to the width of the diffraction band.

Sec. 23. (a) The wave function of a particle has the form

$$\psi(x,\,t) = \int\limits_{-\infty}^{\infty} C\,(p)\,e^{-i\,(Et-px)/h}\,dp$$

where the amplitude $C\,(p)$ is assumed to be other than zero only close to a certain $p = p_0$, and $E = p^2/(2m)$. This wave function describes a so-called *wave packet*. Find the propagation velocity of the maximum value of the amplitude of the wave function, or in other words, the velocity of the wave packet.

(b) In the previous exercise, taking $C\,(p) = e^{-(p-p_\bullet)^2/[2(\Delta p)^2]}$, where $p_0 \gg \Delta p$, show that the minimum spatial width of the wave packet satisfies the inequality $\Delta x > [ht/(2m)]^{1/2}$. Here, the width is determined from

$$|\,\Delta\psi\,|^2 \sim e^{-(x-\varphi t)^2/[2\,(\Delta x)^2]}$$

according to the amplitude of the wave function.

Sec. 24. (a) Develop the spherical functions Y_1^0, Y_2^0, Y_3^0, Y_1^1, Y_2^1, Y_2^2, Y_3^1, Y_3^2, Y_3^3, and verify their orthogonality.

(b) Find the commutator of $\hat{M}_x^2$ and $\hat{M}_y^2$.

(c) Find the commutator of $\hat{p}_x$ and $1/r$.

Sec. 25. (a) Compute $\langle \cos^2 \vartheta \rangle$ in states with wave functions Y_1^0, Y_1^1, which must first be normalized.

(b) The azimuthal dependence of the wave function of a particle is other than zero and constant for $0 \leqslant \varphi \leqslant \pi$, and equal to zero for $\pi < \varphi < 2\pi$. Expand it in a set of eigenfunctions of the orbital angular momentum projection and determine the probability of a certain eigenvalue of the angular momentum projection.

(c) Let ϑ, φ be the polar angle and azimuth of a point with respect to axis z, and θ and χ be the same with respect to axis x. Using the fact that $\cos \vartheta = \sin \theta \cos \chi$, $\cos \theta = \sin \vartheta \sin \varphi$, expand the spherical functions $Y_1^k(\vartheta,\,\varphi)$ in the function $Y_1^{k'}(\theta,\,\chi)$.

Sec. 26. (a) Express $e^{-\hat{r}^2/(2r_0^2)}$ (where r_0 is a constant quantity) in the momentum representation.

(b) Show that the operator $\hat{A}$ of a finite shift, whose operation on a function is defined as $\hat{A}\psi\,(x) = \psi\,(x + a)$, has in momentum representation the form $\hat{A} = e^{-iap_x/h}$.

Sec. 27. The probability of the appearance of a certain eigenvalue of energy E_n in a system is $w_n = e^{-\beta E_n}$, where $\beta > 0$. Find the mean value of the quantity λ in coordinate representation.

Sec. 28. (a) Find the energy spectrum of a particle in a potential "tunnel" of infinite length and constant rectangular cross section. The walls are impermeable for the particle.

(b) Find the energy spectrum of a charged particle in a constant, homogeneous magnetic field. Represent the vector potential of the field in the form $A_x = 0$, $A_y = H_x$, $A_z = 0$ (the *Landau gauge*).

Sec. 29. (a) Consider the one-dimensional motion of a particle in a potential field of the form

$$U = D\,(1 + e^{-2ax} - 2e^{-ax}), \qquad \text{where} \qquad -\infty < x < \infty$$

and show that the discrete energy spectrum is given by the formula

$$E_n = ha\left(\frac{D}{2m}\right)^{1/2}\left(n + \frac{1}{2}\right) + \frac{h^2 a^2}{2m}\left(n + \frac{1}{2}\right)^2$$

with n limited from above. Plot the result on a graph.

Hint. Substitute $e^{-ax} \equiv y$, $\psi = f/y^{1/2}$ and compare with Eq. (29.32), then make use of Eq. (29.37).

(b) Consider the motion of a particle in the field of an attracting centre the potential of which is $U = -\alpha/r^2$. Show that for small values of l there is no solution that near zero is represented by a real and positive power of r, which in classical mechanics corresponds to falling onto a centre.

(c) What are the values of angular momenta obtained in the addition of three angular momenta respectively equal to 1, 2, and 3?

Sec. 30. (a) Given an operator $\hat{\tau} = 3\,(\hat{\sigma}_1 \cdot \mathbf{n})\,(\hat{\sigma}_2 \cdot \mathbf{n}) - (\hat{\sigma}_1 \cdot \hat{\sigma}_2)$, where $\hat{\sigma}_1$ and $\hat{\sigma}_2$ are the Pauli operators for two particles, and $\mathbf{n}$ is a unit vector along the line joining the particles. Show that the eigenvalues of $\hat{\tau}$ are equal to -4, 0, and 2, the latter being two-fold degenerate.

(b) Explain why the most general form of the density matrix in the spin-variable space of one particle of half-integral spin is

$$\rho = a + b\,(\mathbf{n} \cdot \hat{\sigma})$$

where $\mathbf{n}$ is a unit vector. Find $\langle \sigma_x \rangle$.

Sec. 31. (a) Determine, in the quasi-classical approximation, the energy levels in a potential field $U(x) = D\,(1 + e^{-2ax} - 2e^{-ax})$. Explain the obtained result with the help of a mathematical analog with Kepler's problem (compare with problem (a) to Section 29).

(b) Show that in the quasi-classical approximation Kepler's problem yields correct values of the energy levels if the angular momentum square is taken to be $(l + 1/2)^2$ rather than $l\,(l + 1)$.

(c) In the quasi-classical approximation, find the probability of a particle of energy E penetrating a potential barrier of the form $U = U_0 - \alpha x^2$, $E \ll U_0$, $\alpha > 0$.

Sec. 32. (a) A particle is placed in a spherical potential well, such that $U = -\,|\,U_0\,|$ for $r \leqslant r_0$, and $U = 0$ for $r > r_0$. The depth of the well varies by a constant quantity $U \to -(U_0 + \delta U_0)$. Consider the change in the energy eigenvalue in the first approximation of perturbation theory and compare with the exact formula for the energy of bound states.

(b) The nucleus of a deuterium atom possesses quadrupole moment q. Find the splitting of the levels of the $2p$ state of the electron.

(c) A system has two energy levels the separation of which is $E_1 - E_0 \equiv$ $\equiv \hbar\omega_0$. It is subject to a perturbation $V(t, x)$ which depends upon time according to the law $V(t, x) = \int\limits_{-\infty}^{+\infty} V(\omega, x) \cos \omega t \, d\omega$. At the initial time the amplitude of the zero state $c_0 = 1$, so that the amplitude of the first state $c_1 = 0$. Determine the amplitude of the first state after a sufficiently long time.

Hint. A "sufficiently long time" is a time interval for which we can use the relationship

$$\int\limits_0^\infty f(\omega) \frac{e^{i(\omega - \omega_0)t} - 1}{\omega - \omega_0} \, d\omega = \pi i f(\omega_0)$$

which is developed similarly to formula (32.39).

Sec. 33. (a) What power of the atomic number is the mean distance of the electron from the nucleus calculated by the Thomas-Fermi method proportional to?

(b) Using only parity considerations, obtain the selection rule for matrix elements of the coordinate with respect to the quantum number l, similar to the way it was developed for the vector $\boldsymbol{\gamma}$ with respect to J (Exercise 4, Section 33).

Sec. 34. What are the rotational states of an oxygen molecule O_2^{16} and its heavy isotope O_2^{17}? (The nuclear spin of O^{17} is determined by one neutron above the occupied shells.)

Sec. 35. (a) In the Born approximation, find the differential cross section of elastic scattering of electrically charged particles of dipole moment d.

Hint. Use the limiting transition from a system of two charges at finite distance from each other to a dipole.

(b) Find the partial scattering cross section of a particle in the field of a repulsive centre of potential $U = \alpha/r^2$ if it is known that the asymptotic solution of the equation $y'' + (2\lambda/x) \, y' + k^2y = 0$ is finite for $x = 0$, and that for $x \to \infty$, $y = x^{-\lambda} \cos(kx - \pi\lambda/2)$.

Sec. 36. (a) Taking the quantized motion of an oscillating dipole, calculate the radiation intensity in a transition from the first excited state to the ground state, assuming the dipole moment to be proportional to the oscillator coordinate. Compare with the corresponding classical formula.

(b) Calculate the radiation intensity of a hydrogen atom in a transition from the $2p$ to the $1s$ state.

(c) What is the multipole order of the transitions

$$^4S^g_{1/2} \longrightarrow {}^4S^u_{3/2}, \qquad {}^3D^g_1 \to {}^3P^g_1, \qquad 2P_{3/2} \longrightarrow 2P_{1/2}?$$

Sec. 37. (a) Separate the variables in the Dirac equation for a central field, assuming the first two functions to be dependent only upon the radius.

Hint. Choose the first pair of wave functions in the form 0, ψ_2 and ψ_1, 0; seek the angular dependence of the second pair from the form of the spherical functions for $l = 1$.

(b) Show that if the Dirac wave function is subjected to the transformation $\psi' = C_m\psi \equiv (1/\sqrt{2})\,(\hat{\alpha}_y + \hat{\beta})\,\psi$, the Dirac equation will involve only real coefficients. All the α, β matrices are assumed to have been selected in accordance with the equations of Section 37.

(c) Show that the wave functions of an electron and a positron are of opposite parity.

(d) On the basis of the result of the preceding problem, show that an electron-positron system in a state with total spin 1 can annihilate only by disintegrating into three quanta.

Hint. Take into account that an electron-positron interchange leaves the Hamiltonian invariant only if the sign of the amplitude of the electromagnetic field is simultaneously changed.

INDEX

TO THE READER

Mir Publishers welcome your comments on the content, translation, and design of this book.

We would also be pleased to receive any proposals your care to make about our future publications.

Our address is:
USSR, 129820, Moscow I-110, GSP
Pervy Rizhsky Pereulok, 2
Mir Publishers

NEW BOOKS FROM MIR PUBLISHERS

Fundamentals of Quantum Mechanics
by V. Fock, Mem. USSR Acad. Sci.

This unique course in quantum mechanics was written by an outstanding physics theoretician, who contributed much to the founding and development of the quantum theory. Many of the chapters bear the stamp of his scientific investigations. The first Russian edition appeared in 1932 and for some years it was the only Soviet book on the quantum theory. It has retained its significance to this day, but has long been a bibliographical rarity. For the present edition the author has revised and greatly augmented the book, including his latest findings in quantum mechanics. He elaborates on the epistemological bases of the science, devoting several sections to specific problems that add to an understanding of the theory. Two of the new chapters consider Pauli's theory of the electron and the many-electron problem in its application to the atomic theory.

Contents. **The Basic Concepts of Ouantum Mechanics.** The Physical and Epistemological Bases of Quantum Mechanics. The Mathematical Language of Quantum Mechanics. Quantum Mechanical Operators. The Probabilistic Interpretation of Quantum Mechanics. **Schrödinger's Theory.** The Schrödinger Equation. The Harmonic Oscillator. Perturbation Theory. Radiation, the Theory of Dispersion, and the Law of Decay. An Electron in a Central Field. The Coulomb Field. **Pauli's Theory of the Electron. The Many-Electron Problem of Quantum Mechanics and the Structure of Atoms. Dirac's Theory of the Electron.** The Dirac Equation. The Use of the Dirac Equation in Physical Problems. On the Theory of Positrons.

Molecular Physics

by A Kikoin D Sc and I Kikoin, Mem. USSR Acad. Sci.

The authors acquaint first-year students specializing in the subject with the fundamentals of the molecular structure of matter on the basis of the kinetic theory. They give the essential elements of thermodynamics, statistical physics, and the quantum theory needed to understand the subject. Great attention is devoted in the book to the experimental methods used in physics.

Contents. Kinetic Theory of Ideal Gases. Kinetic Theory of Heat. Law of Conservation of Energy. Collisions of Molecules and Transport Phenomena. Physical Phenomena in Rarefied Gases (a Vacuum). Real Gases. The Van der Waals Equation. Elements of Theormodynamics. Properties of Liquids. Low Temperatures. Solids.

VOLUME 1 OF THIS *COURSE OF THEORETICAL PHYSICS* DEALS WITH THE FUNDAMENTAL LAWS OF PHYSICS. THE TEXT LUCIDLY PRESENTS FOR STUDENTS AND WORKERS IN THEORETICAL PHYSICS THE FUNDAMENTAL PRINCIPLES UNDERLYING THE FINDINGS OF EXPERIMENTAL PHYSICS. IT GIVES A UNIFIED PRESENTATION OF CLASSICAL MECHANICS, ELECTRODYNAMICS, AND QUANTUM MECHANICS AND PROVIDES AN EXCELLENT FOUNDATION FOR THE STUDY OF MORE ADVANCED TOPICS IN ATOMIC, MOLECULAR, AND SOLID-STATE PHYSICS. *FUNDAMENTAL LAWS* CAN BE READ BY STUDENTS WHO HAVE HAD COURSES IN INTRODUCTORY PHYSICS, ELECTRICITY, AND MAGNETISM.

ALEXANDER S. KOMPANEYETS
(1914-1974)

PROFESSOR ALEXANDER SOLOMONOVICH
KOMPANEYETS WAS A LEADING SOVIET
THEORETICAL PHYSICIST. FROM 1946
UNTIL HIS UNTIMELY DEATH HE
WORKED AT THE INSTITUTE OF CHEM-
ICAL PHYSICS OF THE USSR ACADEMY
OF SCIENCES, CONTRIBUTING, AMONG
OTHER THINGS, TO THE DEVELOPMENT
OF NUCLEAR ENERGY IN THE SOVIET
UNION IN ALL ITS ASPECTS.

THEORETICAL MECHANICS: A SHORT COURSE

by S. Targ

This short course of theoretical mechanics is designed for students of higher and secondary technical schools. It deals with the basic methods of theoretical mechanics and their spheres of application. It also briefly examines motion in a gravitational field, the motion of bodies with variable mass (e.g., rockets), the elementary theory of gyroscopic motion, and other topics. These questions are of such importance today that no course of mechanics, even a short one, can neglect them altogether. Considerable space—more than one-third of the total volume of the book—is given to examples and worked problems. They were chosen with an eye to ensure a clear comprehension of the relevant mechanical phenomena, and they embrace all the main types of problems and methods of problem solution. The problem solutions contain instructions designed to assist the student in his independent work on the course. In this respect, the book should prove useful to students from many fields interested in advancing their knowledge of the subject.
525 pages

LECTURES IN ANALYTICAL MECHANICS

by F. Gantmacher

This volume of lectures is concerned with the techniques of analytical mechanics and their application to Lyapunov's theory of stability, the theory of oscillations, and the dynamics of solids. Elements of modern frequency methods are presented, along with classical problems of the theory of oscillations. The lectures provide a substantial foundation for studying the special theory of relativity, quantum mechanics, and other fields of theoretical physics. They are designed for undergraduates in the mechanico-mathematical and physics departments of universities, and also for research engineers and other specialists who feel a need to extend their knowledge in mechanics.
264 pages

HANDBOOK OF PHYSICS

by B. Yavorsky and A. Detlaf

This handbook defines the basic concepts of physics, concisely formulates physical laws and the essence of the phenomena described. All branches of classical and modern physics are covered. The six main parts of the book include: Physical Basis of Classical Mechanics, Fundamentals of Thermodynamics and Molecular Physics, Fundamentals of Fluid Mechanics, Electricity and Magnetism, Wave Phenomena, and Atomic and Nuclear Physics. The handbook is intended for use by engineers, technicians, students of universities and engineering institutes, post-graduate students, teachers of high schools, lecturers of institutes, and technical writers. It can be expediently used by anyone interested in the field.
965 pages

KOMPANEYETS

A COURSE OF

THEORETICAL PHYSICS

1